



MOHEBI &
ASSOCIATES

COMPUTATIONAL + EXPERIMENTAL NEUROSCIENCE

MODEL. EXPERIMENT. REFINE.

Biological reinforcement learning and *dopamine*

CS 839 – Fall 2026

Ali Mohebi, PhD

Department of Psychology,
Neuroscience Training Program,
University of Wisconsin–Madison

Learning objectives



By the end, you should be able to:

- 01** Distinguish learning predictive cues from learning actions
- 02** Describe how neurons communicate and connections change
- 03** Connect dopamine prediction errors to RL algorithms
- 04** Compare action learning, discounting and distributional RL

Lecture outline



- 1 Learning in psychology**
What changes with experience?
- 2 Biological foundations**
What are neurons, synapses and dopamine?
- 3 Prediction errors and plasticity**
How does an outcome change a prediction and a connection?
- 4 Learning to choose**
How do predictions guide actions?
- 5 Rewards over time**
How does delay change value?
- 6 Distributions of future reward**
What is lost when we predict only an average?

Learning becomes an experimental science



Late 1800s

Measure how experience changes behavior

**Ebbinghaus
1885**

Practice and relearning reveal how memory changes over time.

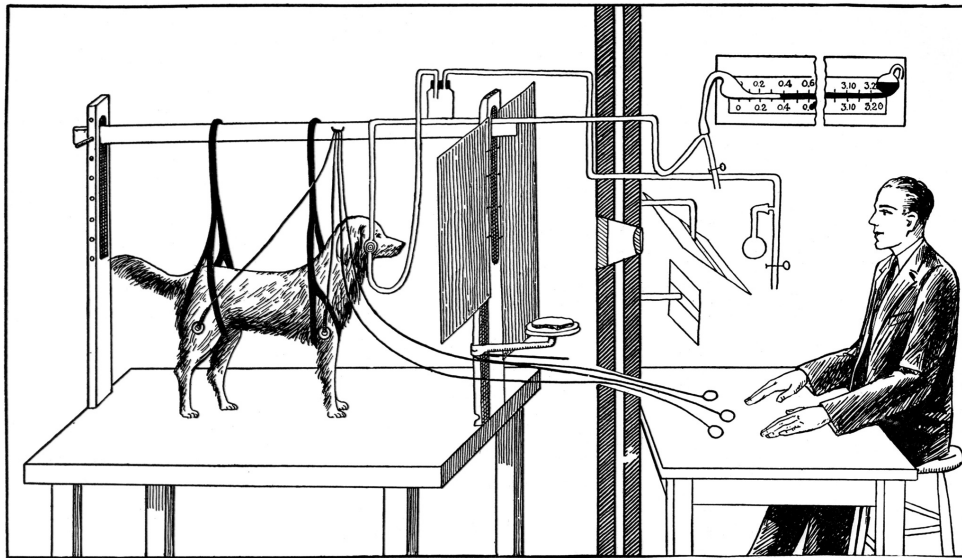
**Animal learning
1890s onward**

Control the experience, measure the response, and compare explanations.

A theory of learning must explain measurable changes caused by experience.

Ebbinghaus (1885); Pavlov (1927); Thorndike (1898/1911)

Pavlov: learning that a cue predicts food



Training

A sound precedes food.

Later test

The sound alone elicits salivation.

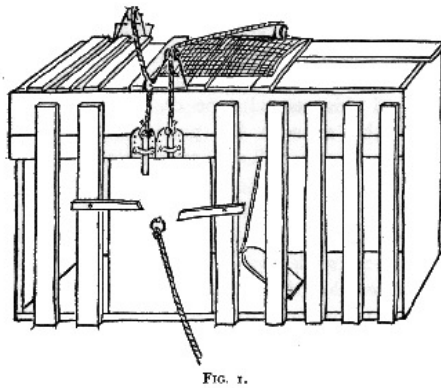
Food delivery does not depend on the dog's response.

Classical conditioning studies how cues come to predict outcomes.

Thorndike: learning through consequences

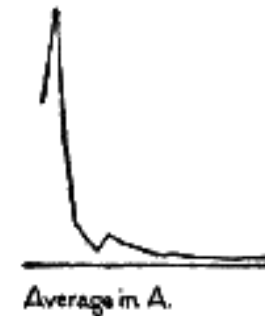


Puzzle box K



An effective action opens the door and gives access to food.

Learning curve for box A



Horizontal: repeated trials
Vertical: time to escape

The law of effect links an action's consequences to its future probability.

Thorndike (1898/1911), *Animal Intelligence*, Figs. 1-2

Classical and operant conditioning



Procedure	What predicts the outcome?	Example
Classical / Pavlovian	A cue precedes an outcome	Sound, then food
Instrumental / operant	An action changes the outcome	Lever press earns food

Skinner (1938): measure how consequences change the rate of an action.

Prediction and action learning can operate together in the same task.

Pavlov (1927); Skinner (1938); Sutton & Barto (2018), ch. 14

Tolman: learning can precede reward

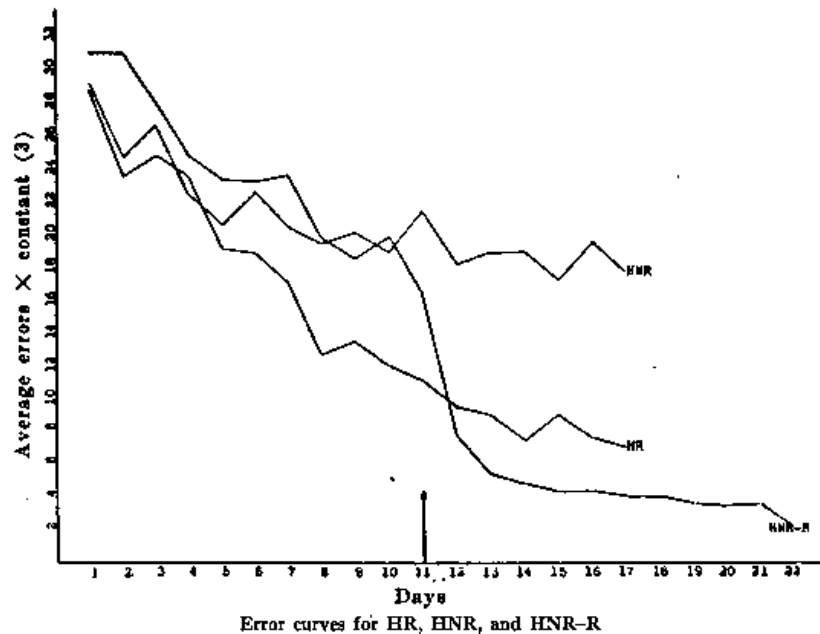


FIG. 6

(From E. C. Tolman and C. H. Honzik, Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 19, p. 267.)

Reward throughout

HR

No reward

HNR

Reward begins on day 11

HNR-R

Observed performance depends on both what was learned and current motivation.

Tolman (1948), Fig. 6, reporting Tolman & Honzik (1930)

What changes during learning?



Perspective	What experience changes
Association	Connections among cues, outcomes or responses
Reinforcement	The future probability of an action
Cognitive representation	Knowledge of the environment and its relationships
Error correction	A prediction, according to its mismatch with an outcome

These are overlapping explanations, not mutually exclusive stages.

Selected perspectives relevant to RL; Sutton & Barto (2018), ch. 14

Marr's three levels of explanation



Level	Question	Biological RL example
Computational	What problem, and why?	Predict future reward to guide action
Algorithmic	What information and update?	Represent a prediction and correct errors
Implementational	What physical system?	Identify the cells and connections

A match at one level does not establish all three.

The computational problem: predicting return



Marr: computational level



$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots$$

$$V(s) = E[G_t \mid s_t = s, \text{policy } \pi]$$

Reward is an outcome. Value predicts future return.

Learning from errors in expectations



Marr: Algorithmic level

Learn by comparing what you expected with what actually happened.

Simple learning rules

- 1 Better than expected?
Increase the tendency to repeat that action.
- 2 Worse than expected?
Decrease the tendency to repeat that action.
- 3 Exactly as expected?
No prediction-error update is needed.

The canonical hypothesis: dopamine is a teaching signal
Prediction error = reward – expectation



(500) Days of Summer
Marc Webb, 2009



The learning rule: correcting a prediction

Marr: Algorithmic level

$$\delta = r - V \quad V' = V + \alpha\delta$$

Each example starts at $V = 0.5$, with learning rate $\alpha = 0.2$.

Outcome	Reward r	Error δ	New value V'
Better than expected	1	+0.5	0.6
As expected	0.5	0	0.5
Worse than expected	0	-0.5	0.4

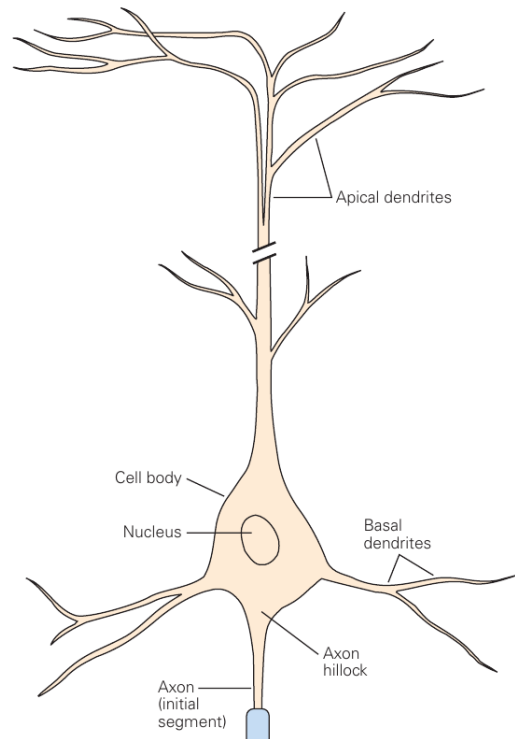
An outcome changes value according to how much it differs from the prediction.

Illustrative one-step learning rule; Rescorla & Wagner (1972); Sutton & Barto (2018)

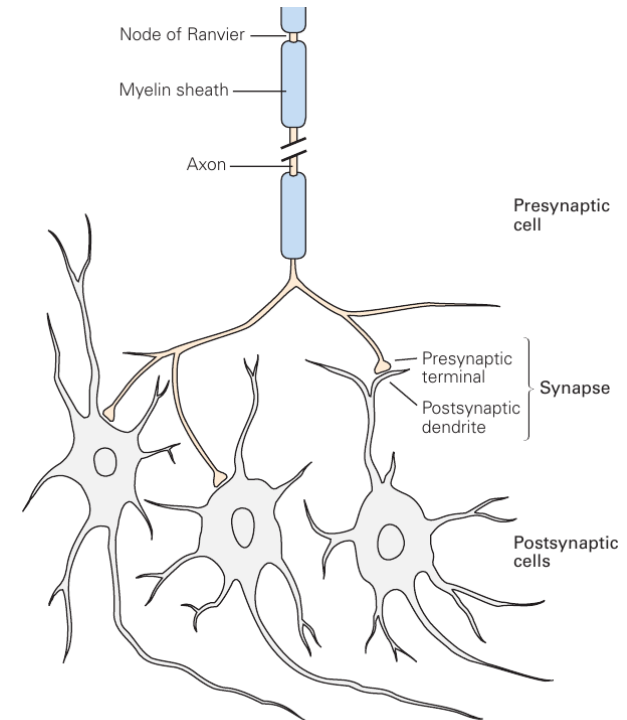
The structure of a neuron



Receiving inputs



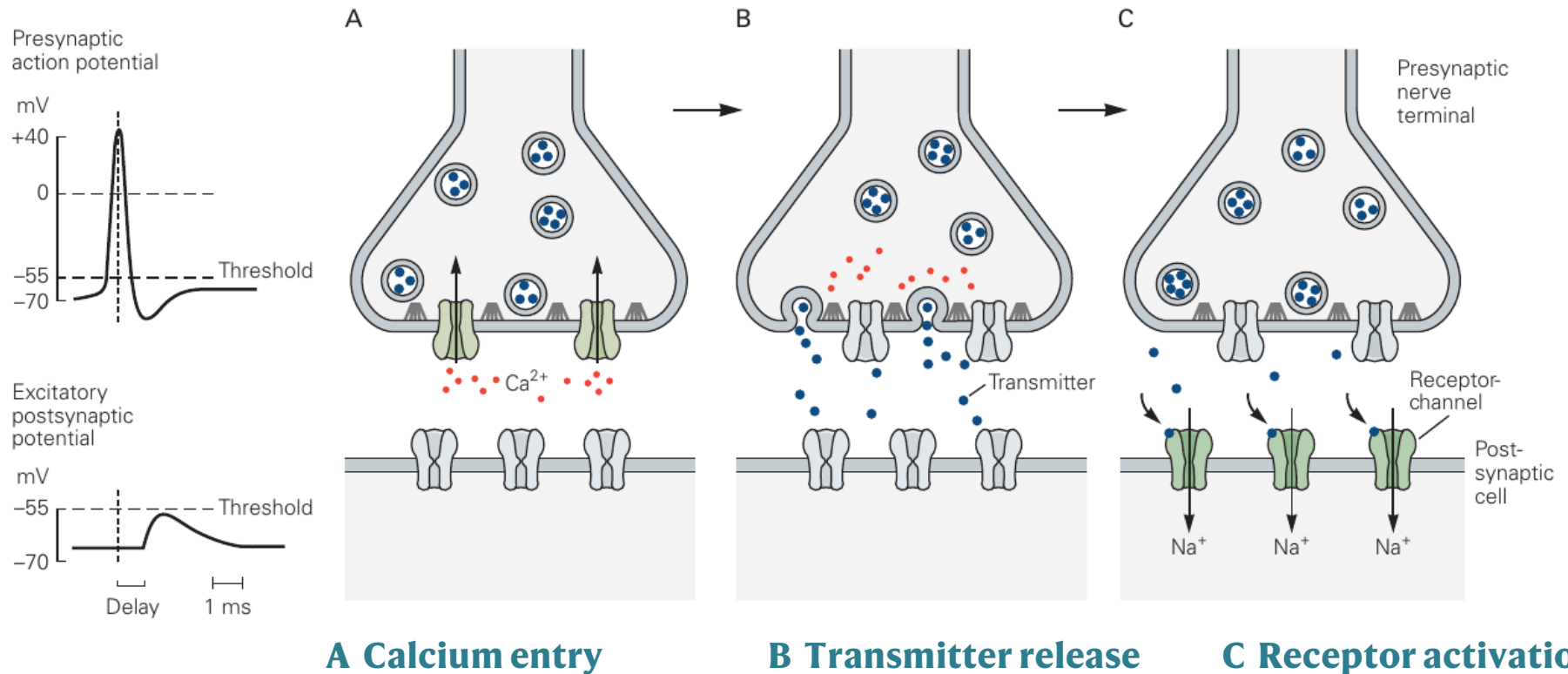
Sending signals to other cells



A spike is a brief electrical pulse. Firing rate counts spikes per second.

Kandel et al., Principles of Neural Science, 6th ed. (2021), Fig. 3-1, p. 57. Enlarged in two portions.

Chemical transmission at a synapse



Dopamine as a neuromodulator



Neuromodulation changes the state of a neuron or network, altering its response to subsequent inputs.

The same input can produce a different response.

Activity now

Dopamine can change excitability and the influence of other inputs.

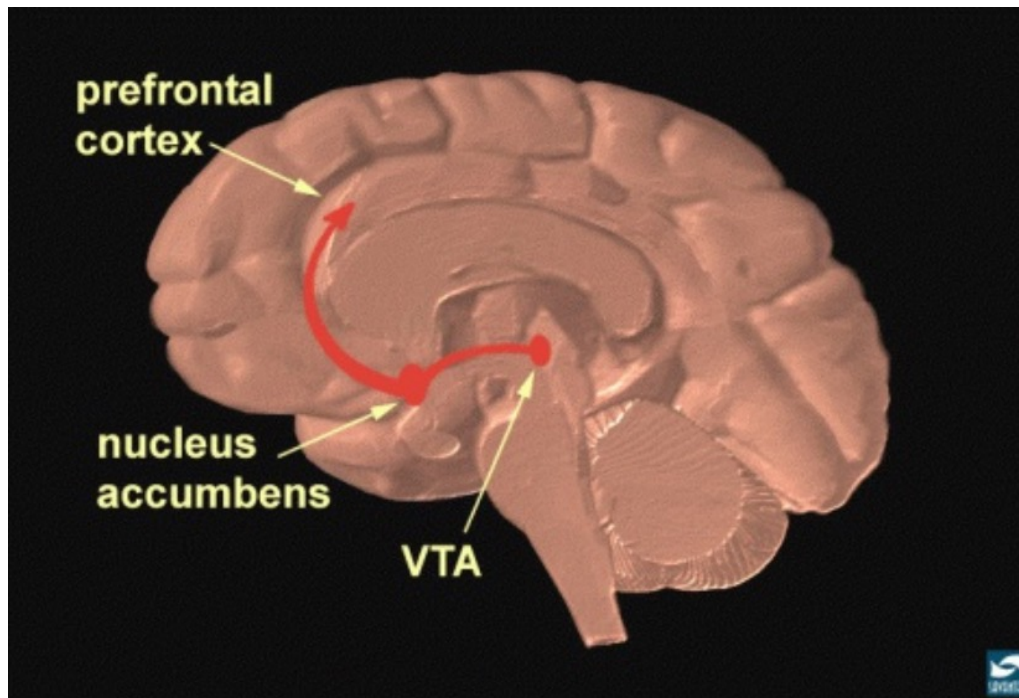
Learning from experience

Dopamine can change how active synapses undergo plasticity.

Its effect depends on the receptors, target cells and ongoing activity.

Definition paraphrased from Picciotto, Higley & Mineur, Neuron (2012). Dopamine: Tritsch & Sabatini (2012).

Dopamine sources and their projections



VTA shown here. The substantia nigra is another major source.

Dopamine neurons

Cell bodies cluster in a few nuclei. Their axons branch widely.

Glutamatergic neurons

Cell bodies occur across the brain, with local and long-range connections.

Anatomy image: NIDA (2007). Tritsch & Sabatini (2012), Matsuda et al. (2009), Kandel et al. (2021).

Schultz, Dayan & Montague, 1997



1997

A Neural Substrate of Prediction and Reward

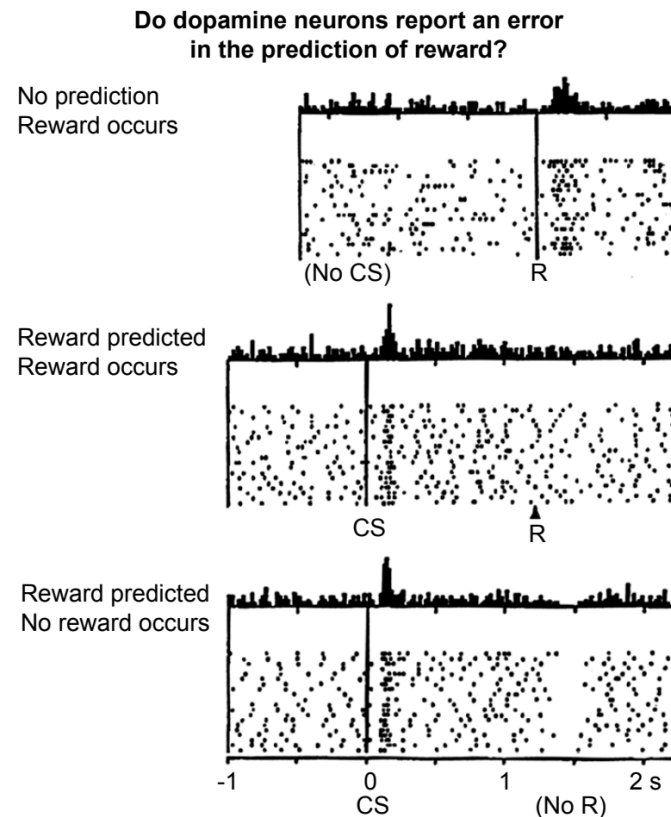
Wolfram Schultz, Peter Dayan, P. Read Montague*

A computational
account
of a biological signal

Conditioning + dopamine physiology + temporal-difference learning

Science 275:1593–1599 (1997), doi:10.1126/science.275.5306.1593

What changes when reward is predicted?



Unexpected reward

A burst at reward

Predicted reward

The burst moves to the cue

Expected reward omitted

A pause when reward was due

Each dot is a spike. Each row is a trial. CS = cue. R = reward.

Schultz, Dayan & Montague, Science (1997), Fig. 1

A prediction error can occur before reward



$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

Reward + discounted next prediction – prior prediction

Event	Reward	Next value	Prior value	TD error
Unexpected cue	0	1	0	+1
Predicted reward	1	0	1	0
Reward omitted	0	0	1	-1

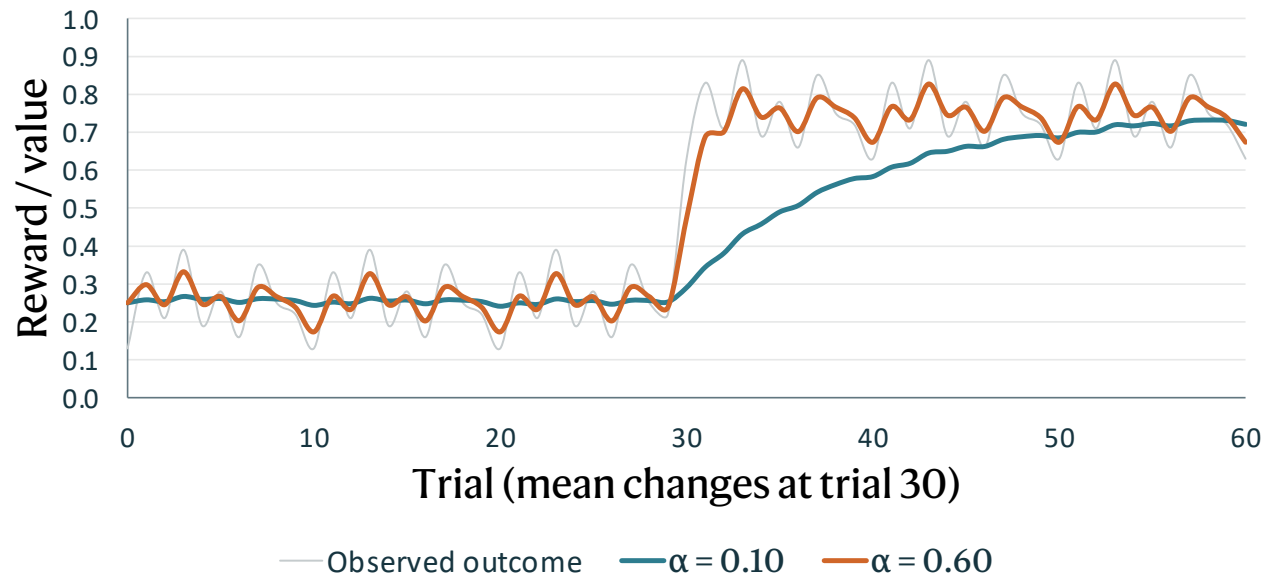
Illustrative transitions: $\gamma = 1$, terminal value = 0, cue arrival unexpected.

Illustrative TD transitions; Sutton & Barto (2018), ch. 14–15

Learning rate sets the weight of new evidence



$$V' = V + \alpha\delta$$



In AI

Choose a rate or schedule using validation tasks. Adapt it when useful.

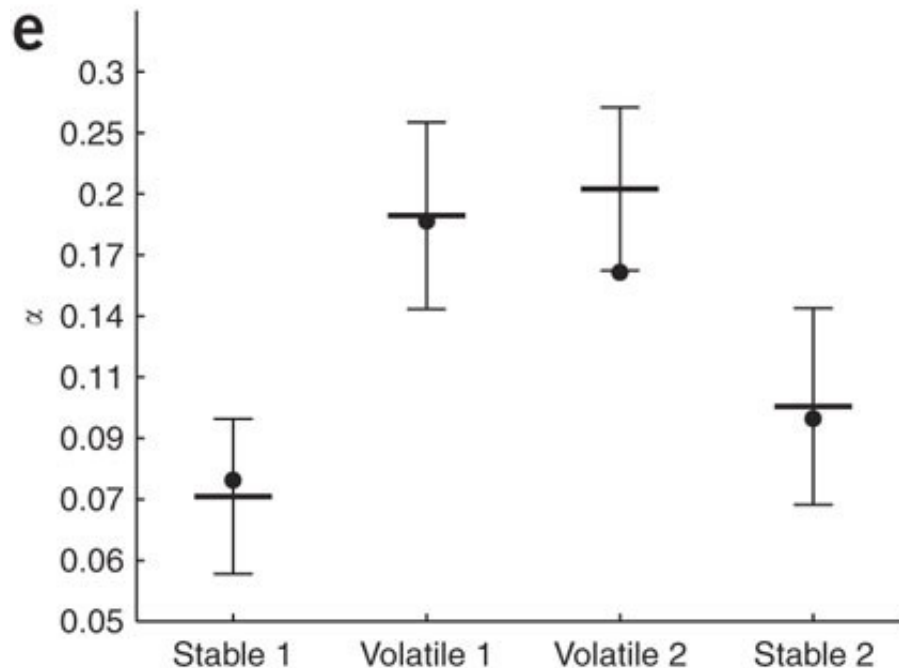
In biology

Infer learning rates from behavior and compare fixed vs. adaptive rules.

Fast updating follows change quickly, but also follows noise.

Illustrative simulation; Sutton & Barto (2018), Ch. 2; Nassar et al. (2010)

People learn faster in volatile environments



Bars: human estimates (mean ± SEM) Dots: Bayesian model

Stable

Outcome probabilities remain constant.

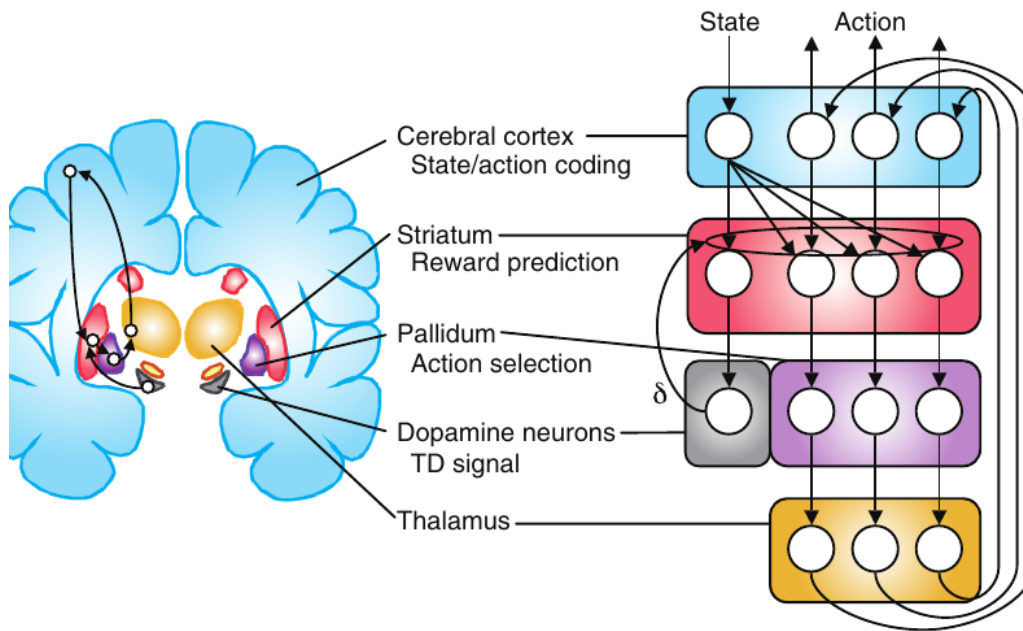
Volatile

Outcome probabilities change during the task.

Learning rate can adapt when the environment changes.

Behrens et al., Nature Neuroscience (2007), Fig. 2e

Cortical input and dopamine meet in the striatum



Marr: implementation

Cortex

Glutamatergic inputs reach striatal neurons.

Dopamine

Modulates their activity and synaptic plasticity.

Corticostriatal connections are candidate sites for learning.

Doya (2008), Fig. 2; hypothetical model

Learning changes synaptic strength



Strength means how much an input changes the receiving cell's response.

Before strengthening

Same input



Response



Small

After strengthening

Same input



Response



Larger

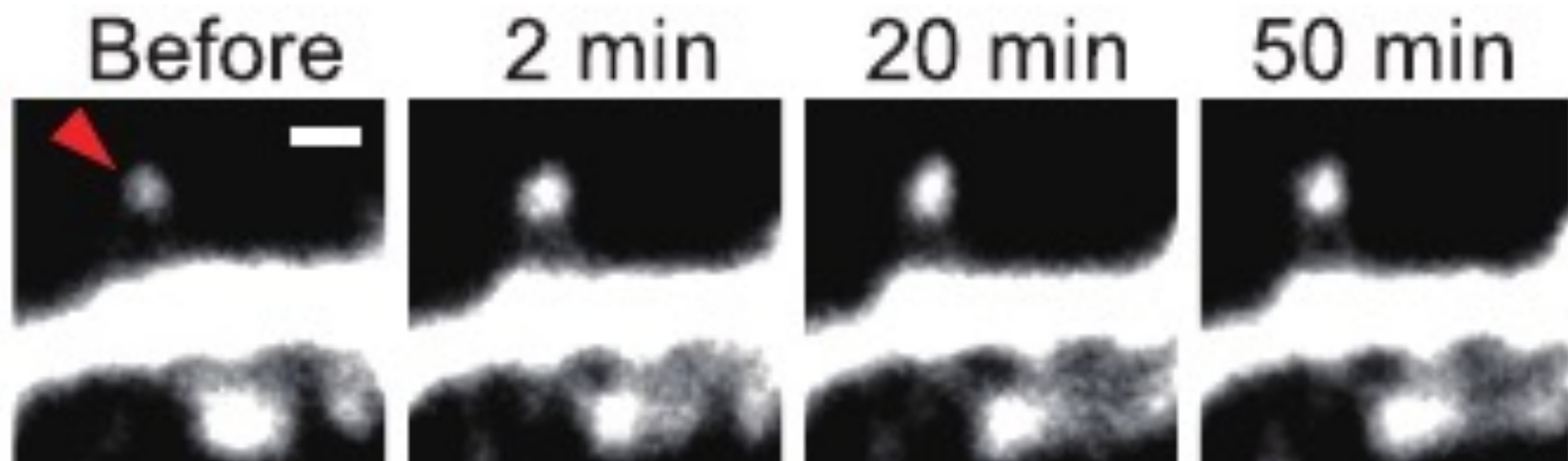
Synaptic plasticity = a lasting change in this influence.

Following one dendritic spine over time



A spine is a small protrusion from a dendrite.

The red arrow marks the spine stimulated in this experiment.



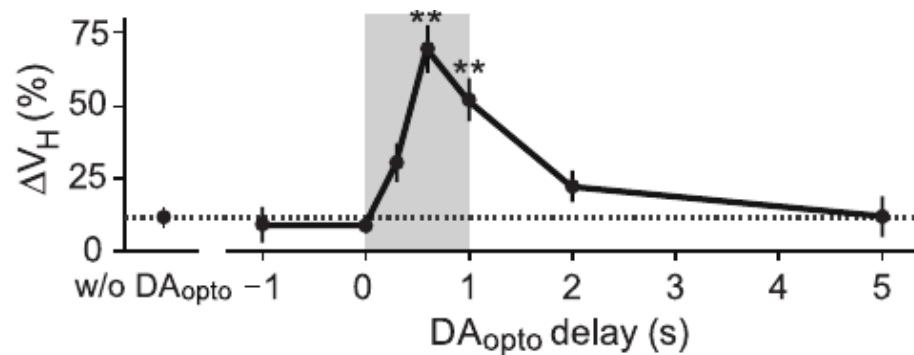
Spine enlargement is a local structural change associated with plasticity.

Yagishita et al., Science (2014), Fig. 1E

Dopamine acts within a narrow time window



Experiment: activate a striatal spine, then vary the timing of dopamine.



Glutamate activates a spine

Dopamine follows ~0.6–1 s later

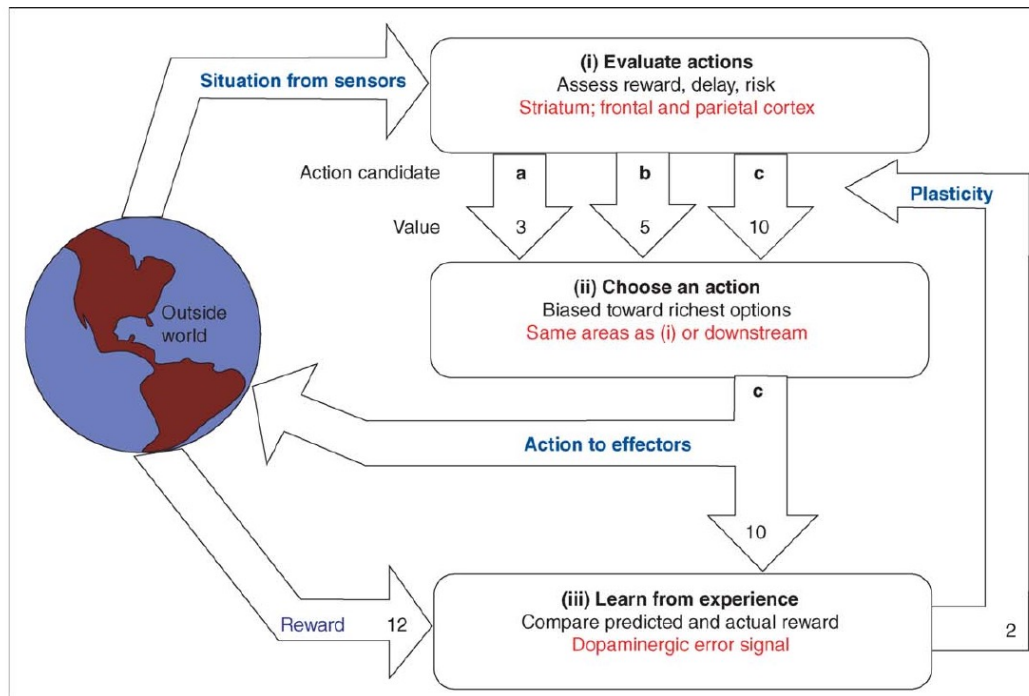
Spine enlargement is strongest

A brief window for local plasticity

Spine growth depends sharply on when dopamine follows synaptic activity.

Yagishita et al. (2014), Fig. 10

Values guide actions and outcomes update values



1 Predict

Estimate outcomes for candidate actions.

2 Choose

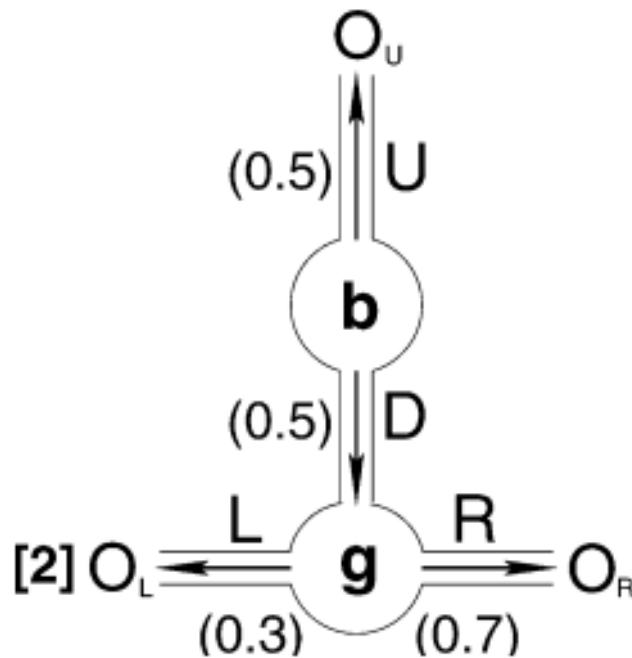
Use value to guide action.

3 Learn

Compare outcome with prediction.

Prediction error closes the loop and improves future choices.

Actor and critic



Critic: predict return

$$V(g) = 0.3 \times 2 + 0.7 \times 0 = 0.6$$

Actor: change the policy

Increase the probability
of the better action

A prediction error can train value estimates and action preferences.

Dayan & Balleine, Neuron (2002), Fig. 1; Dayan, RL 2 (supplied)

Reward and omission can update differently



Start with $V = 0.5$. The outcome is either 1 or 0.

Outcome	Prediction error	Learning rate	Updated value
Reward = 1	$\delta = +0.5$	$\alpha^+ = 0.30$	$V' = 0.65$
Omission = 0	$\delta = -0.5$	$\alpha^- = 0.10$	$V' = 0.45$

These are two separate trials starting from the same prediction.

OpAL splits action learning into Go and NoGo

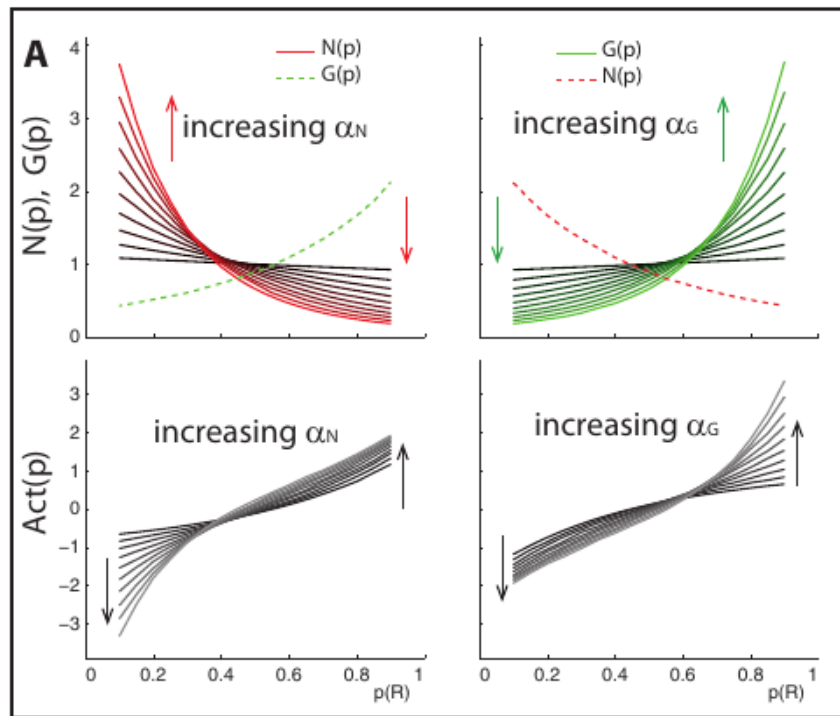


The critic computes a prediction error: $\delta = \text{reward} - \text{expected reward}$

Pathway	Learning update	Contribution to action
Go (G)	$G' = G + \alpha G \cdot G \cdot \delta$	Supports choosing the action
NoGo (N)	$N' = N - \alpha N \cdot N \cdot \delta$	Opposes choosing the action

Choice uses $\text{Act} = \beta G \cdot G - \beta N \cdot N$

Different rates favor good or bad options



Higher α_G

Separates high-value options

Higher α_N

Separates low-value options

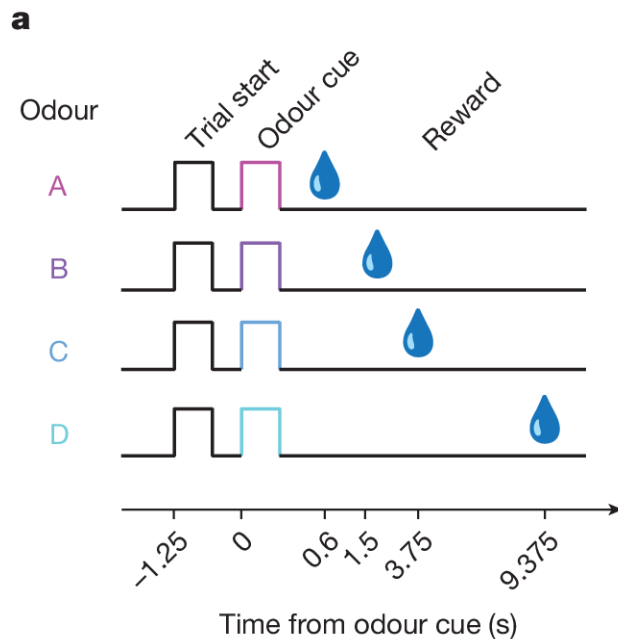
Model simulations

Learning-rate asymmetry changes which actions are easiest to distinguish.

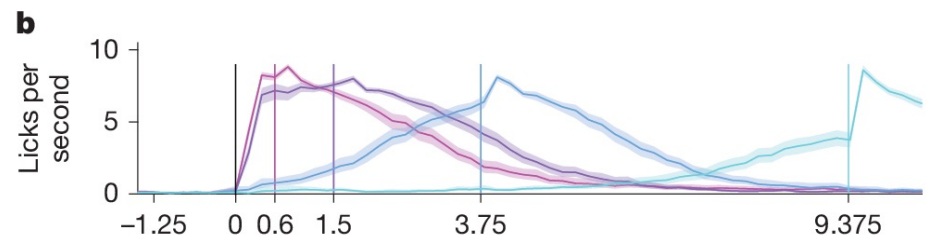
How does reward delay change value?



Same reward, different delays



Mice learn when to expect it



Color identifies the odor and delay
Licking measures anticipation

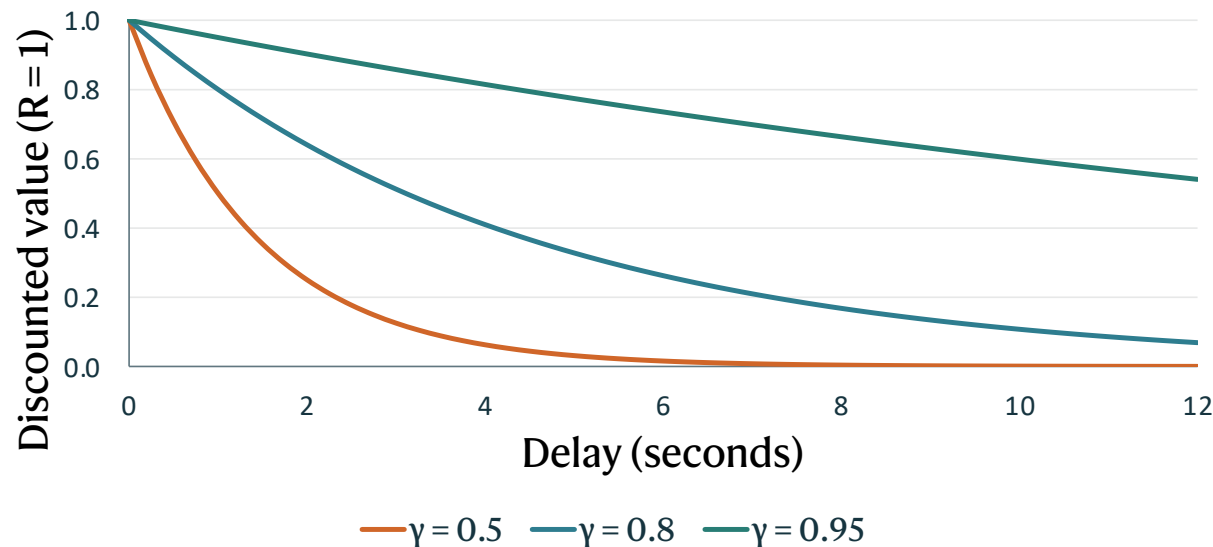
Does the same reward have the same value when it arrives later?

Masset et al., Nature (2025), Fig. 2a,b

Exponential discounting



$$V(\text{delay}) = R\gamma^d = R \exp(-d/\tau)$$



Short horizon
 $\gamma = 0.5$

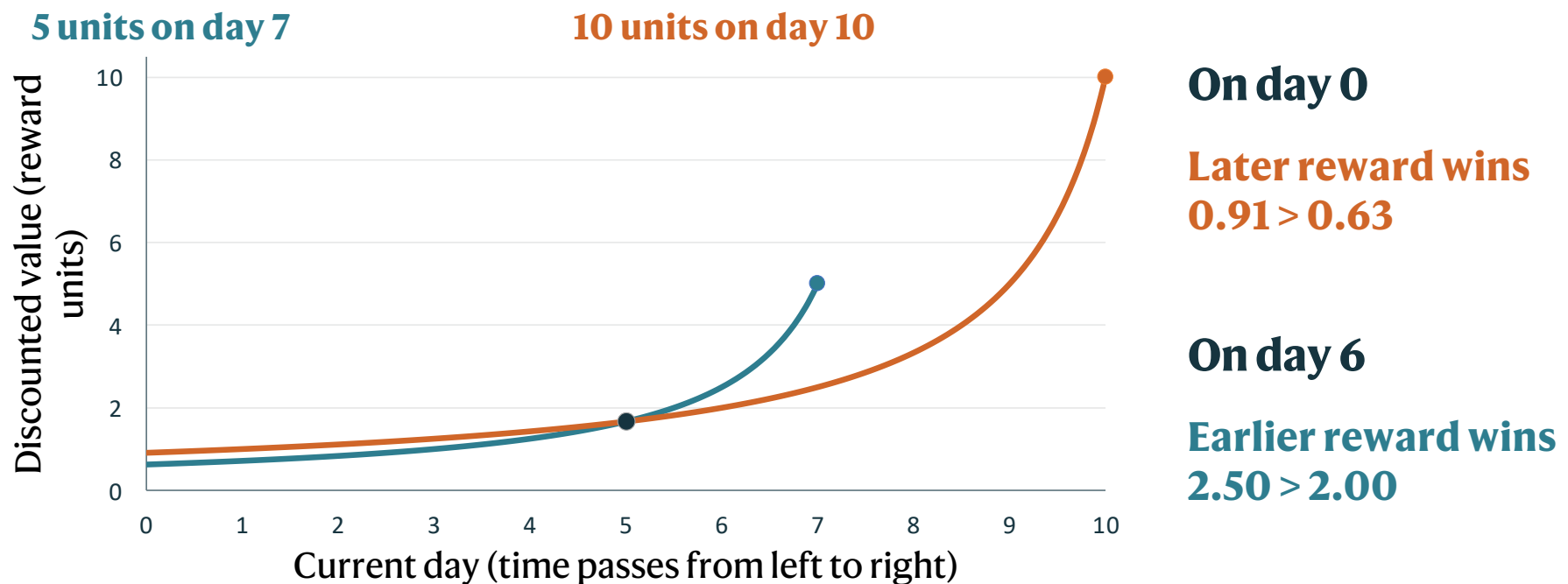
Long horizon
 $\gamma = 0.95$

γ is defined per second
in this example.

Preference can reverse as rewards approach



Same fixed rewards: 5 units on day 7 or 10 units on day 10

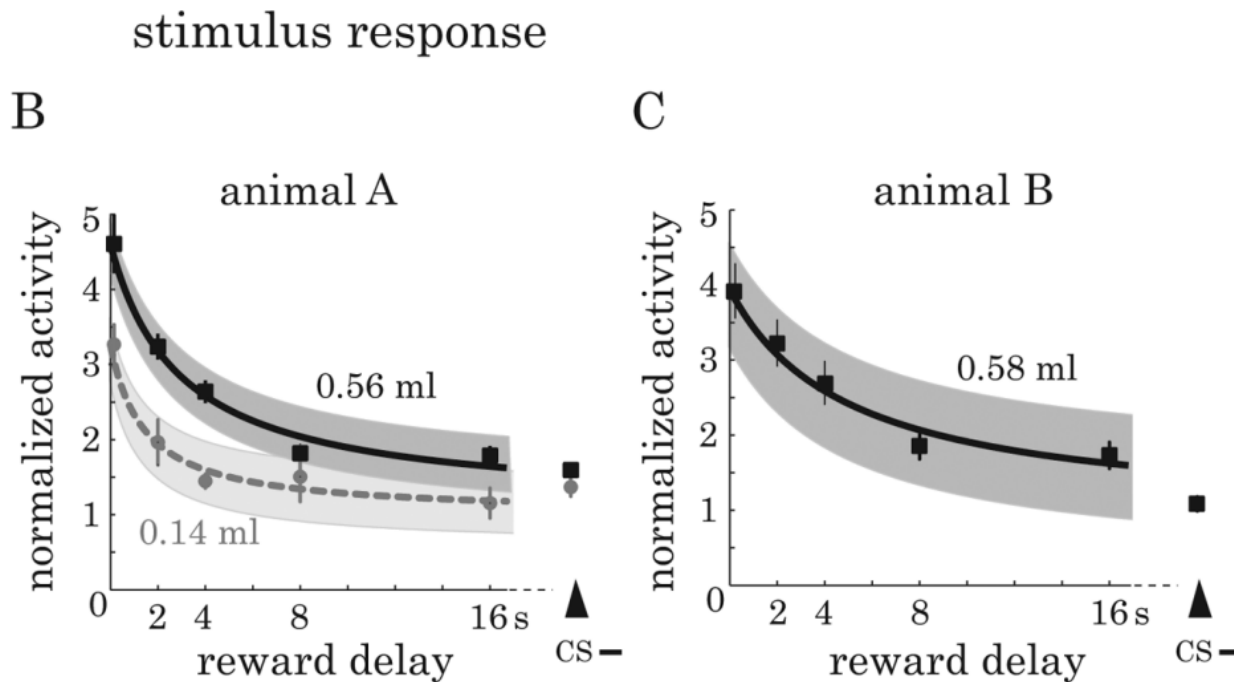


Equal value on day 5. After that, the smaller, sooner reward wins.

$V = R / (1 + kD)$, $k = 1/\text{day}$. $D = \text{days remaining until delivery}$.

Illustrative calculation: $V = R / (1 + kD)$, $k = 1$ per day; see Sutton & Barto, Ch. 15

Dopamine responses discount delayed rewards



Kobayashi & Schultz
varied reward delay in
monkeys

Cue responses decreased
as reward delay increased

Behavior and dopamine
responses showed
temporal discounting

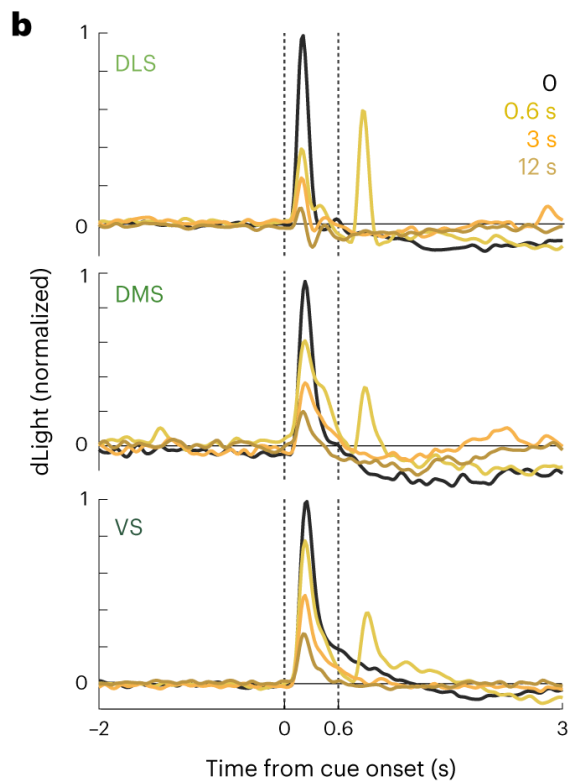
Kobayashi & Schultz (2008), Fig. 6B,C

Kobayashi & Schultz, Journal of Neuroscience (2008)

Exponential vs. hyperbolic dopamine discounting

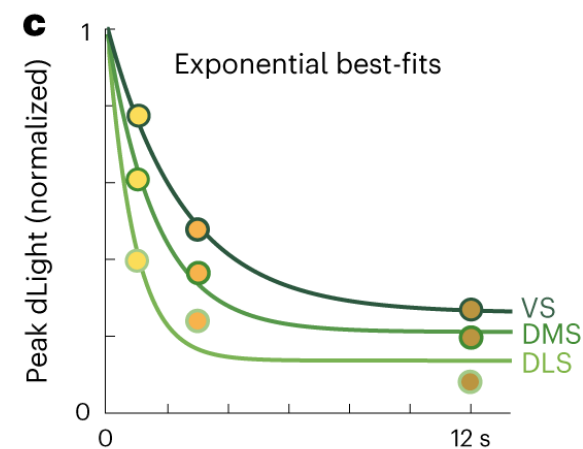


Dopamine responses



Exponential

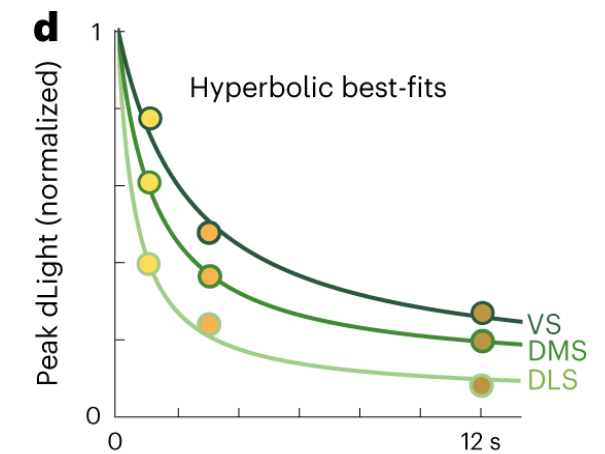
$$\exp(-D/\tau)$$



Same fractional loss
for each extra second

Hyperbolic

$$1 / (1 + D/\tau)$$



Smaller proportional penalty
at longer delays

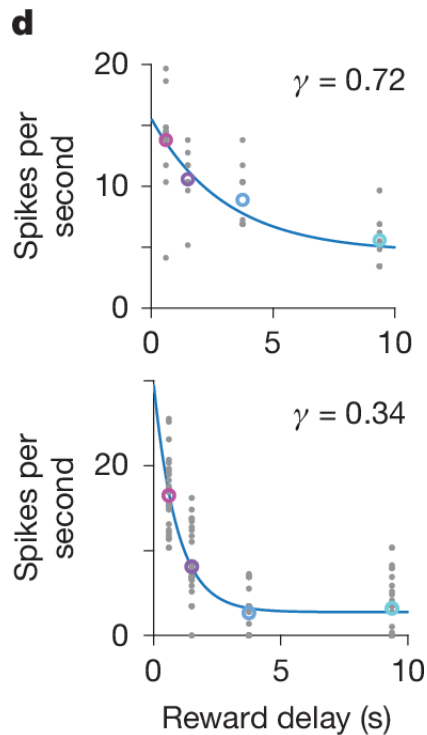
Both fits: DLS discounts fastest, DMS is intermediate, VS slowest.

Mohebi, Wei et al., Nature Neuroscience (2024), Fig. 5b–d. Fits include scale and baseline terms.

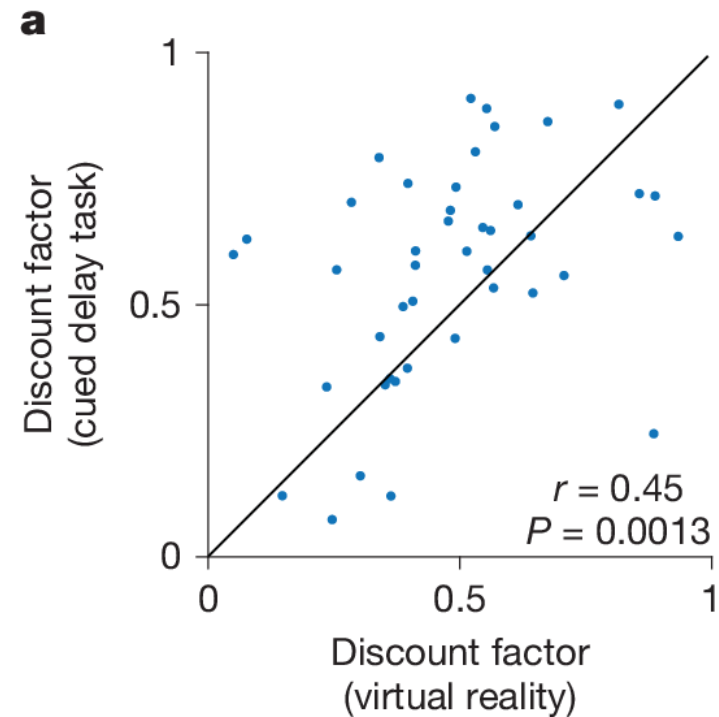
Individual dopamine neurons span time horizons



Two example neurons



Same neuron, two tasks



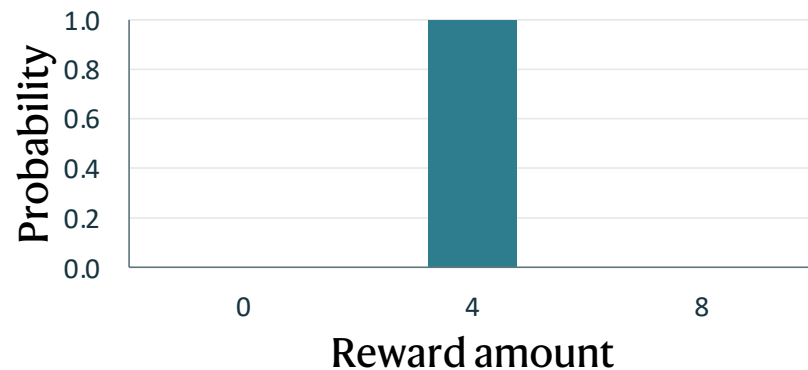
Differences across neurons partly persist across behavioral contexts.

Masset et al., Nature (2025), Fig. 2d and Fig. 4a

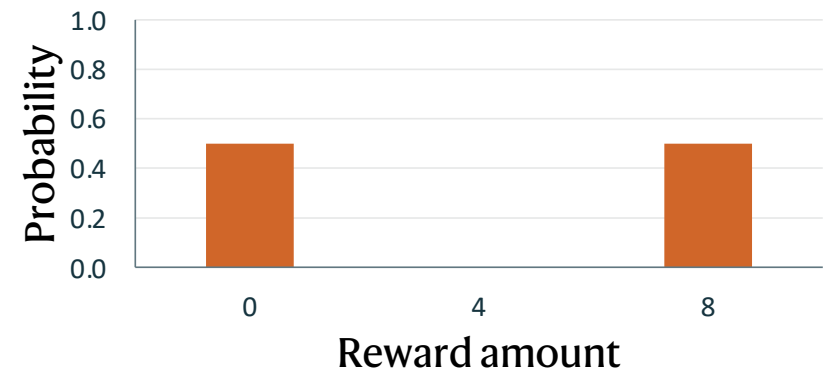
Same average, different chances of success



A: always 4



B: 0 or 8, equally likely



Question	A	B
Expected reward	4	4
Chance of at least 3	100%	50%
Chance of at least 6	0%	50%

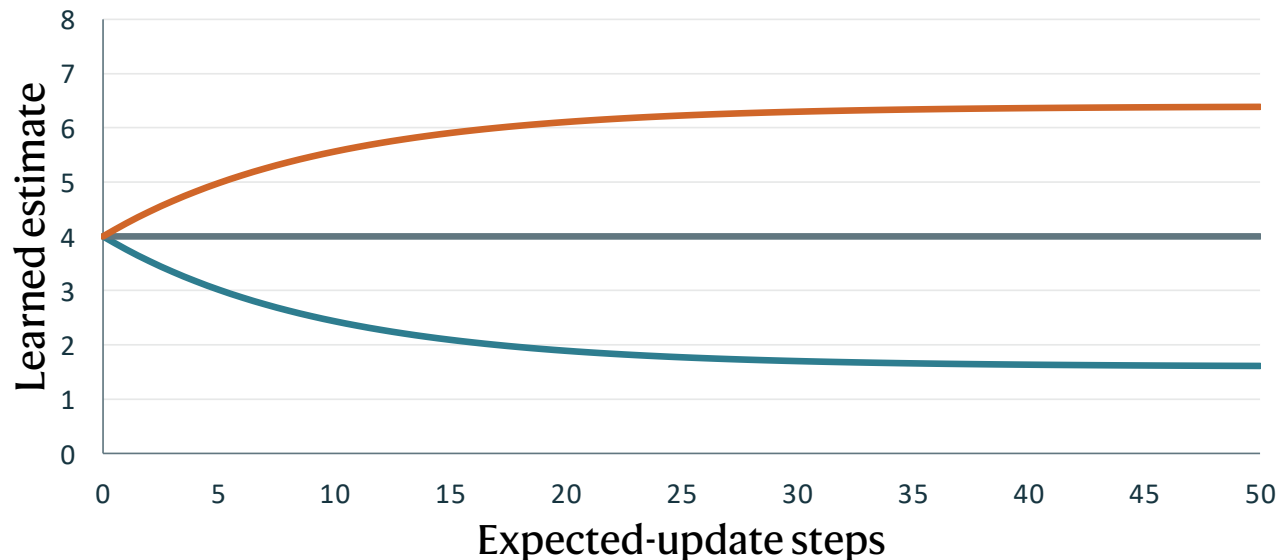
A distribution predicts possible returns and their probabilities.

Illustrative one-step reward distributions

Asymmetric updates learn different predictions



Same outcomes: 0 or 8, equally likely. All estimates start at 4.



More positive weight

$\alpha+ = 0.16, \alpha- = 0.04$

Balance point: 6.4

Equal weights

$\alpha+ = \alpha- = 0.10$

Balance point: 4.0

More negative weight

$\alpha+ = 0.04, \alpha- = 0.16$

Balance point: 1.6

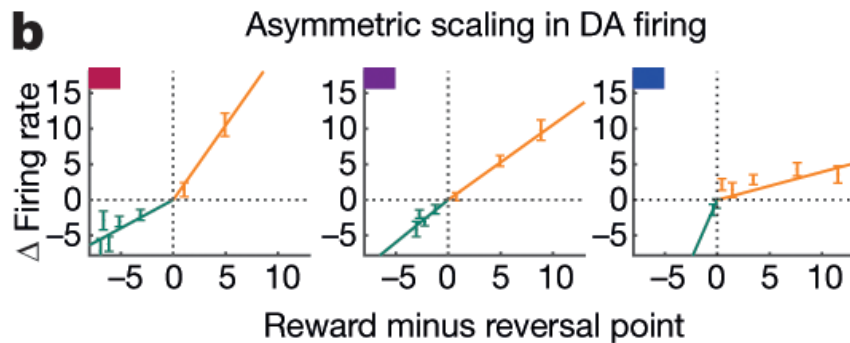
These balance points are expectiles. This is a model prediction to test in neurons.

Illustrative expected-update simulation; Dabney et al. (2020); Rowland et al. (2019)

Dopamine neurons have different value baselines



Measured response curves



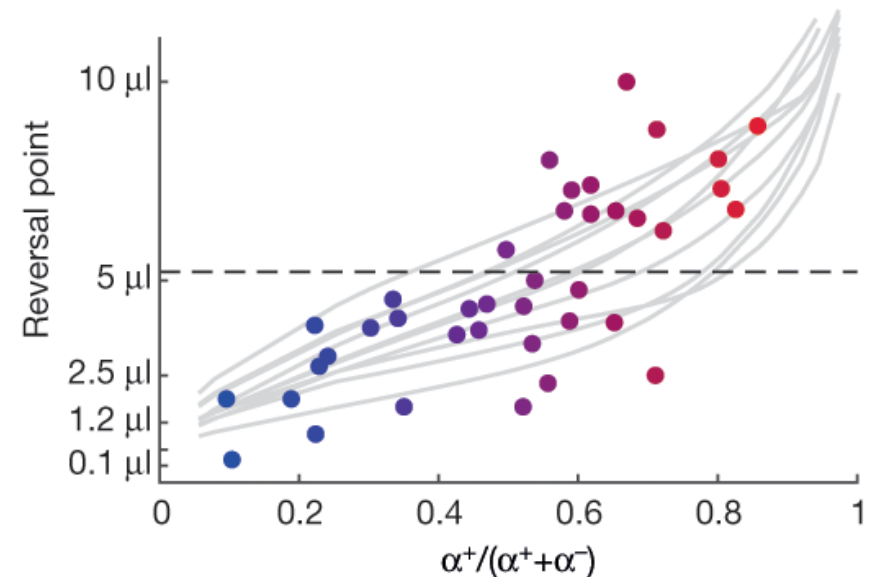
Reversal point: the reward amount where excitation changes to inhibition

Response asymmetry predicts where each neuron changes sign.

Dabney et al., Nature (2020), Fig. 4b,d

Population test

d Asymmetry predicts reversal point

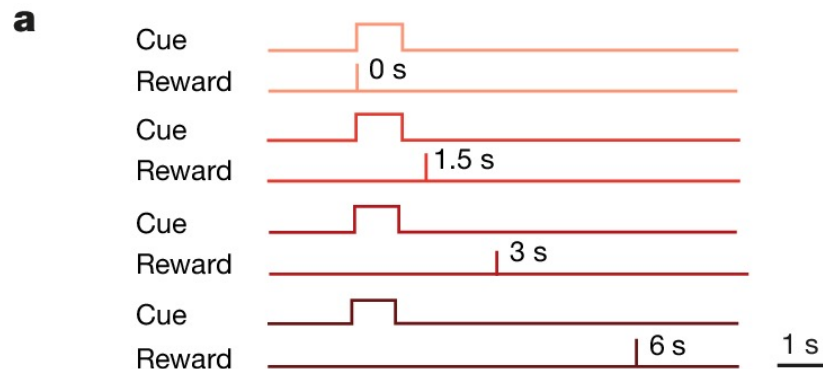


The experiment separates reward amount and delay



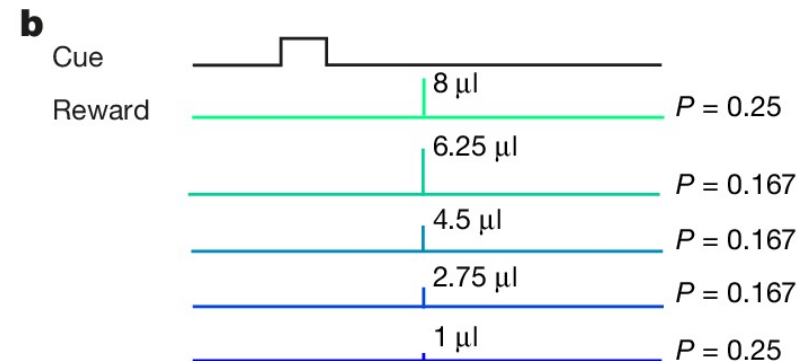
Mice learn five odor cues. What can each cue tell them about the future?

Four cues: same amount, different delays



4.5 μ l after 0, 1.5, 3 or 6 s

One cue: variable amounts, fixed delay



1–8 μ l after 3 s; mean = 4.5 μ l

At 3 s, the fixed and variable cues have the same mean reward.

Decoding asks what the population predicts



Can we reconstruct the possible rewards before they arrive?

1 Measure the population response at the cue

Use 450 ms of firing from neurons with different reward and delay tuning.

2 Combine those responses using a model

Account for each neuron's sensitivity to amount and delay.

3 Reconstruct a distribution, then compare it with the task

The output is a set of possible rewards and their relative likelihoods.

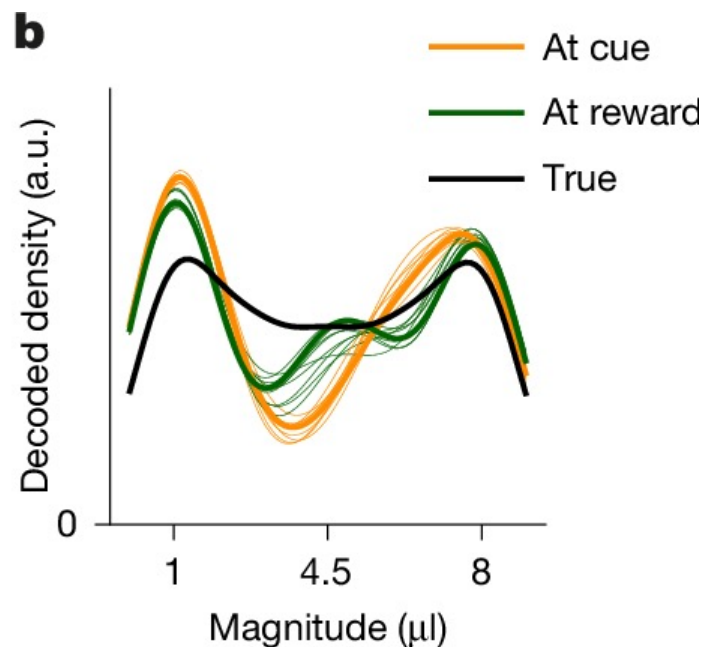
The decoder is an analysis performed by the researchers.

Sousa et al., Nature (2025), Fig. 3d, Fig. 4b,c and Methods

The cue response contains reward-amount information



Focus on the variable cue: reward arrives at 3 s, but its amount varies.



Black: the task's reward distribution

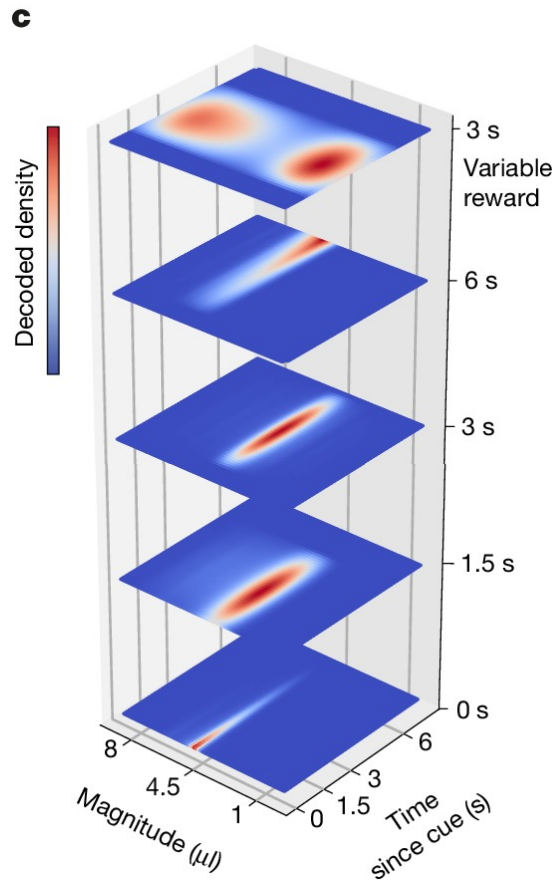
Small and large rewards occur more often.

Orange: reconstructed from cue activity

Green: reconstructed from reward activity

The cue reconstruction captures two peaks, beyond a single mean.

A map can predict both how much and when



Read one plane at a time

Horizontal axes: amount and delay
Warm colors: higher decoded density
Each plane: a different cue condition

Fixed-reward cues

The peak stays near 4.5μ l
and shifts with the cued delay.

Variable-reward cue: top plane

Two likely amounts at the same delay: 3 s.

The joint reconstruction combines amount and time estimates within each cue.

Sousa et al., Nature (2025), Fig. 4c

What the different experiments establish



A useful interpretation separates resemblance, causation and information.

Observation	What it supports	What remains open
Dopamine firing resembles an error	A candidate teaching signal	Which pathways and conditions?
Timed dopamine alters a spine	A causal effect on local plasticity	How does this change learned behavior?
Reward distributions reconstructed from activity	Predictive information is available	Do downstream circuits use it?

Decodable information is evidence for a code, not proof of its readout.

Schultz et al. (1997); Yagishita et al. (2014); Sousa et al. (2025)

A test that holds the average reward constant



Two cues predict rewards after the same 3-second delay.

	Cue A	Cue B
Possible reward	Always 4 units	0 or 8 units, equally likely
Mean reward	4 units	4 units
Timing	3 seconds	3 seconds

At the cue, what should differ across dopamine neurons?

Competing models make different population predictions



Specified model	Prediction at the cue	Useful measurement
One shared mean-value signal	Same expected value for A and B	Mean cue response
A distributional population code	Different patterns across cells, even with equal means	Cell-specific responses and population structure

What biology adds to reinforcement learning



Keep prediction-error learning. Expand what the population predicts.

Across: different reward statistics

Different weighting of positive and negative prediction errors

Depth: different time horizons

Different discount factors retain near and distant outcomes

d Distributional time + magnitude TD learning



A proposed organization of learners,
not an anatomical circuit diagram

Which added prediction would change a decision in your AI application?

Sousa et al., Nature (2025), Fig. 1d (theoretical model)

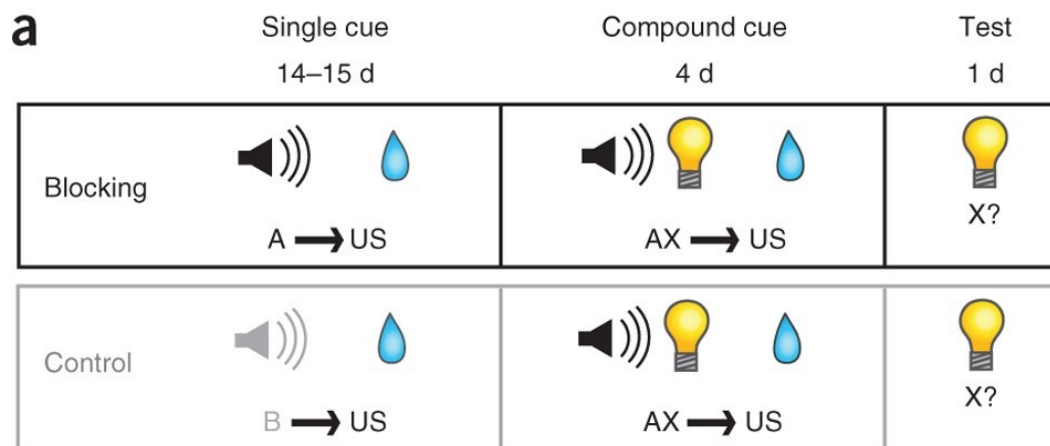
Questions



Which biological constraint would you add to an AI agent?

Which algorithmic variable is hardest to identify from neural data?

Appendix: blocking tests learning from surprise

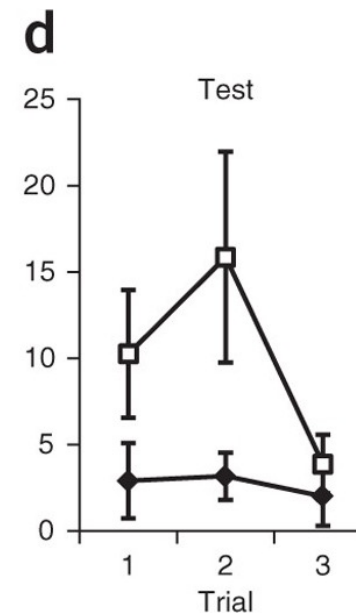


Same light–reward pairing
Different prior prediction

The new cue gains little value when reward is already predicted.

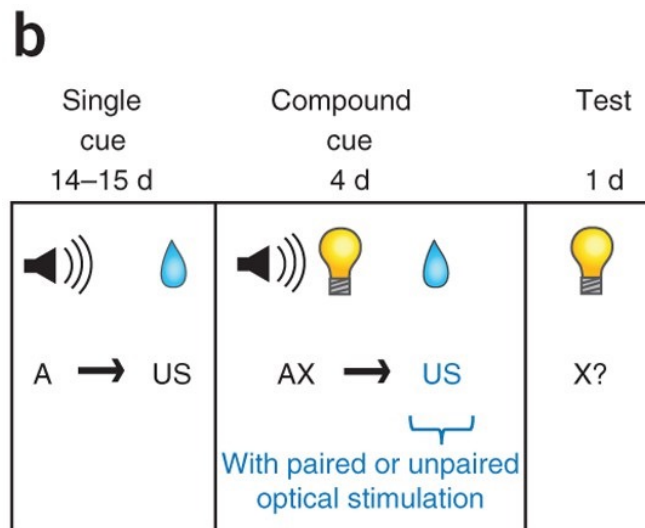
Steinberg et al., Nature Neuroscience (2013), Fig. 1a,d

Reward-port responding (%)



Open squares: control
Filled diamonds: blocking

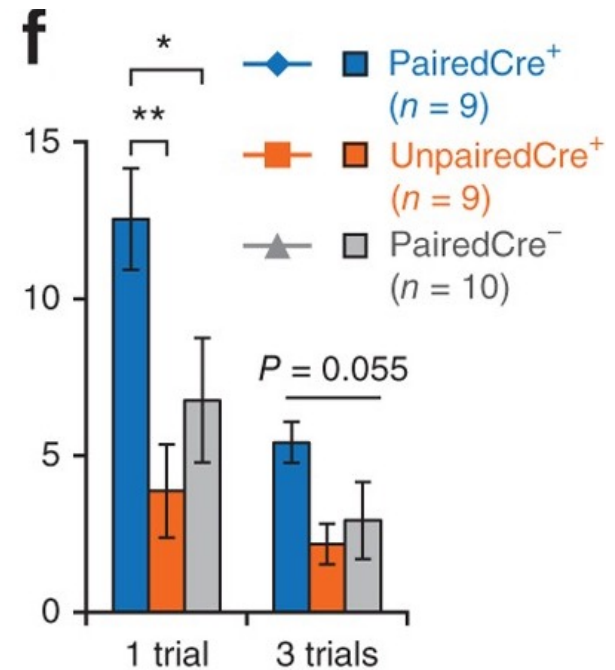
Appendix: dopamine stimulation overcomes blocking



Light activates dopamine cells
at the time of an expected reward

An artificial dopamine signal can overcome blocking.

Reward-port responding (%)

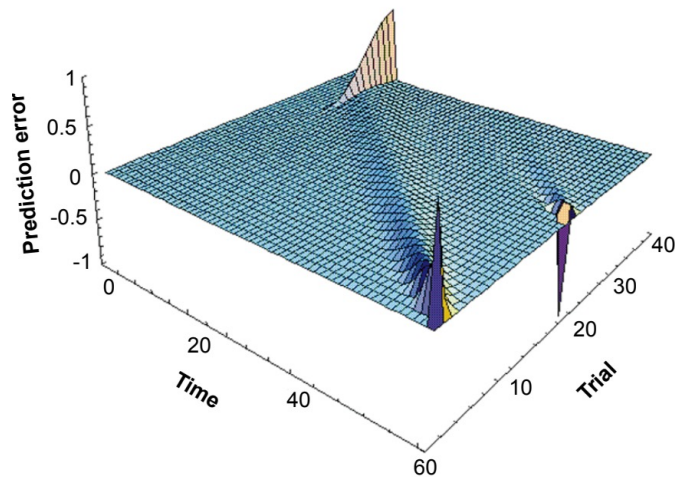


Appendix: TD learning across trials

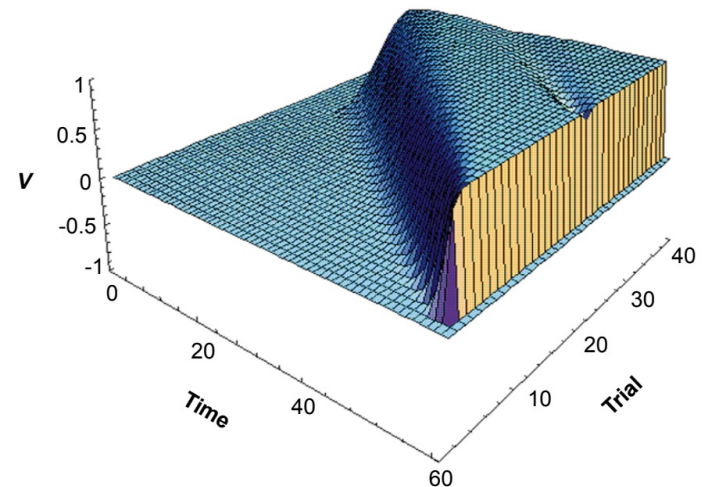


$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

Prediction error



Learned value

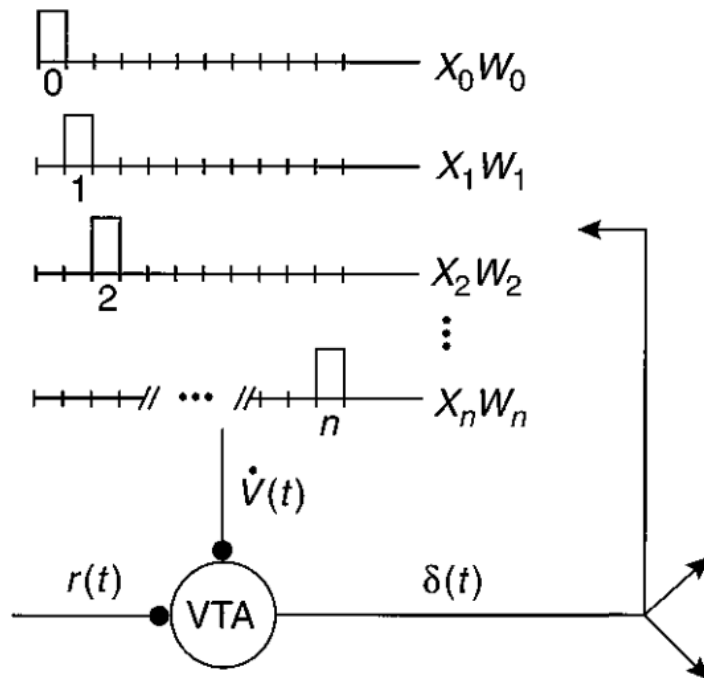


Original model simulations, not neural recordings

The model predicts a transfer of the error to the earliest cue.

Schultz, Dayan & Montague, Science (1997), Fig. 3

Appendix: temporal credit assignment



How does the learner represent elapsed time?

Different features become active after the cue

Eligibility traces preserve recent activity for learning

A learning rule also needs a state representation and a memory of activity.

Schultz et al. (1997), Fig. 2B; Sutton & Barto (2018)

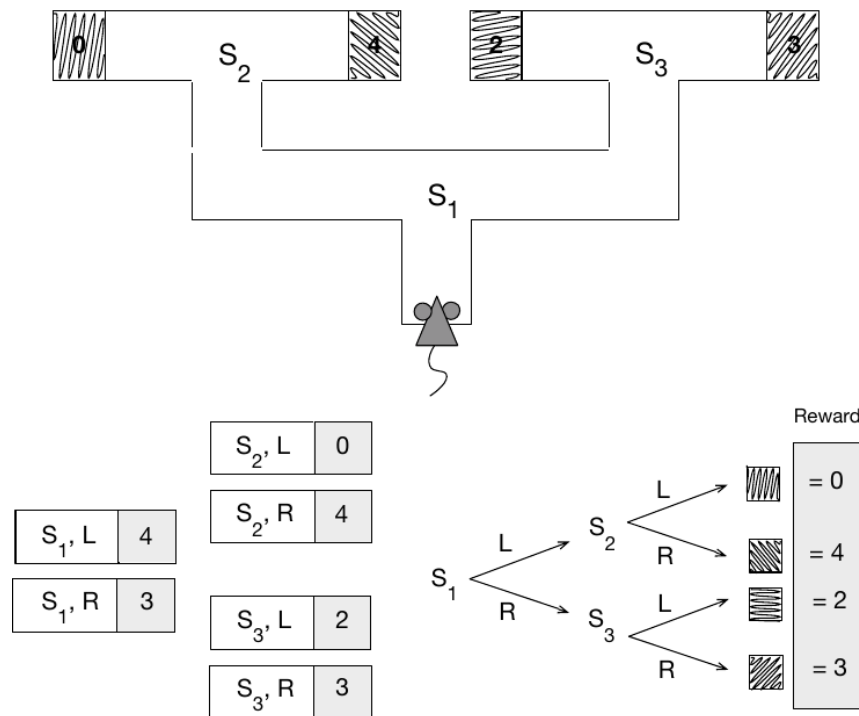
Appendix: second-order conditioning



Phase	What the animal experiences	What can be learned
1	Cue A, then food	A predicts food
2	Cue B, then A (no food)	B predicts a valuable cue
Test	Cue B alone	B can elicit a conditioned response

A reward-only teaching signal misses learning in phase 2.

Appendix: cached values and planning



Model-free

Use learned action values

Model-based

Use transitions and rewards to plan

Does behavior change when an outcome is devalued?

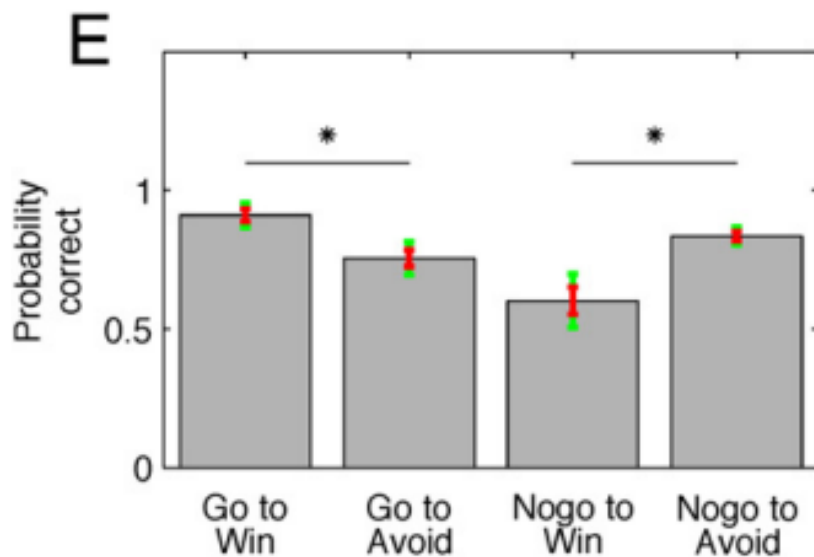
A learned value and a learned world model support different kinds of control.

Sutton & Barto (2018), ch. 14, Fig. 14.9; adapted from Niv, Joel & Dayan (2006)

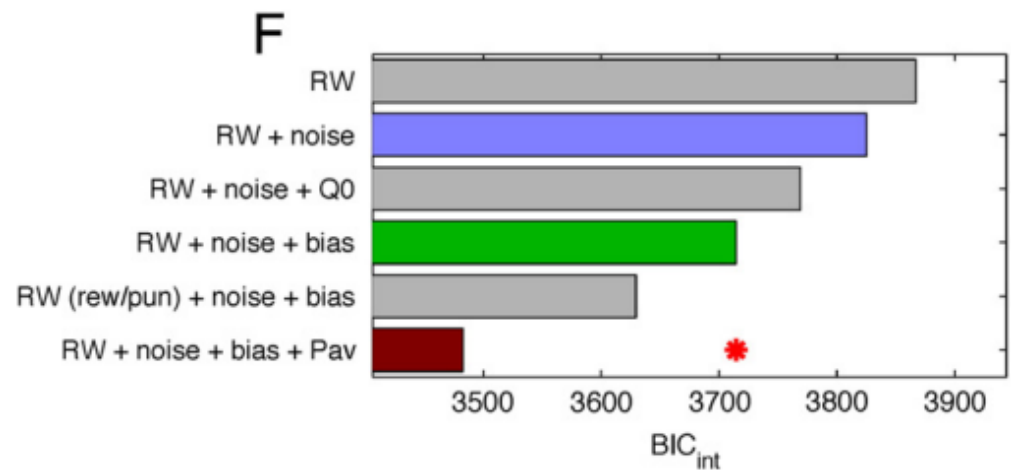
Appendix: Pavlovian biases in action



Observed behavior



Computational explanation



Lower BIC = better fit after a complexity penalty

Harder: withhold a response to win

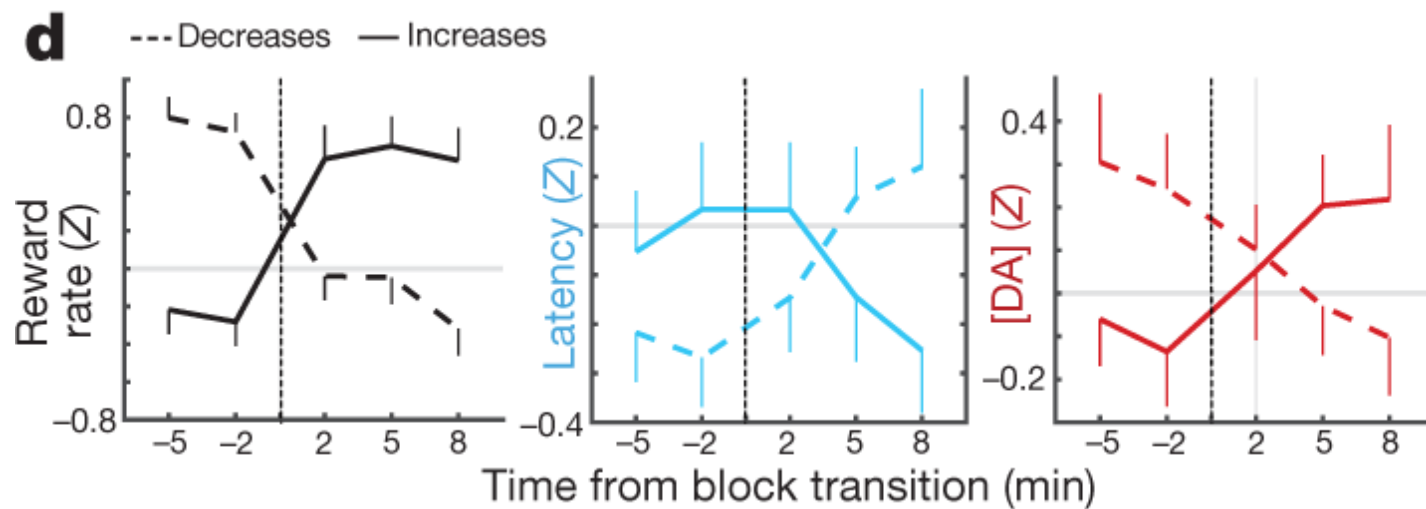
Adding a Pavlovian bias improves the model fit.

Guitart-Masip et al., NeuroImage (2012), Fig. 2E,F

Appendix: dopamine and action vigor



Opportunity cost of waiting $\approx$ average reward rate $\times$ time



Higher reward rate

Shorter initiation latency

More local dopamine

Appendix: one worked TD update



$$r_{t+1} = 0 \quad \gamma = 0.9 \quad V(s_{t+1}) = 0.8$$

$$V(s_t) = 0.5 \quad \alpha = 0.2$$

$$\delta_t = 0.22 \quad \text{updated } V(s_t) = 0.544$$

Illustrative calculation

Appendix: α and γ have different jobs



α : how much does this observation change an estimate?

γ : how much do future rewards contribute to value?

Fast learning can coexist with a long time horizon

Appendix: parameter reference



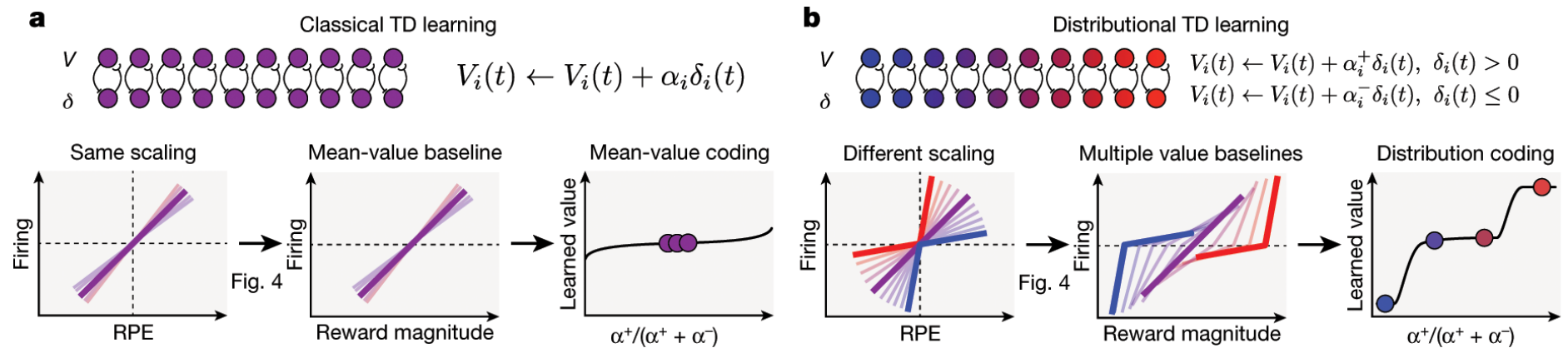
Parameter	Controls	Question
α	Learning rate	How fast does an estimate change?
γ	Future discounting	How far ahead does value look?
λ	Eligibility trace	How long does prior activity remain eligible?
β	Inverse temperature	How strongly do values bias choice?

Appendix: distributional RL formulation



$$Q(s,a) = E[Z(s,a)]$$

$$Z(s,a) \doteq R + \gamma Z(S',A')$$



A population can retain information that disappears in the mean.

Bellemare et al., ICML (2017); Dabney et al., Nature (2020), Fig. 1a,b

Appendix: different roles for asymmetric learning



Model	What learns?	What asymmetry changes
Signed-rate value rule	One value estimate V	Upward versus downward updates
OpAL	Go and NoGo action weights	Approach versus avoidance learning
Distributional expectiles	A population of value estimates	Which part of returns each estimate tracks

Similar mathematics can describe different biological and computational roles.

Appendix: expectile derivation



$$\tau = \alpha^+ / (\alpha^+ + \alpha^-)$$

For $P(R=0) = P(R=8) = 1/2$: $v\tau = 8\tau$

$\tau = 0.2, 0.5, 0.8$ gives $v\tau = 1.6, 4, 6.4$

Appendix: belief-state uncertainty



The learner may infer a hidden state from cues and history

Belief-state values can produce complex dopamine responses

A response pattern alone may not identify the learning algorithm

Appendix: distributional algorithms



Representation	Fixed component	Learned component
C51	Return support locations	Probability at each location
Quantile RL	Cumulative probabilities	Return quantiles
Expectile model	Asymmetry parameters	Asymmetric squared-error estimates

All describe returns, but use different statistics and updates.

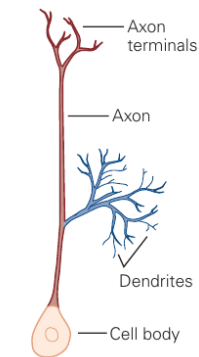
Appendix: neuronal shapes



Unipolar, bipolar, pseudo-unipolar

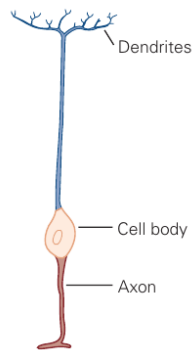
Multipolar neurons

A Unipolar cell



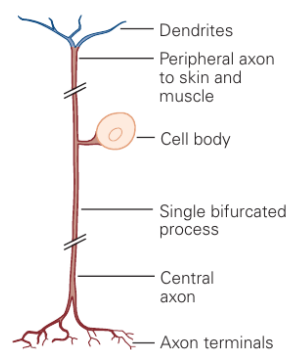
Invertebrate neuron

B Bipolar cell



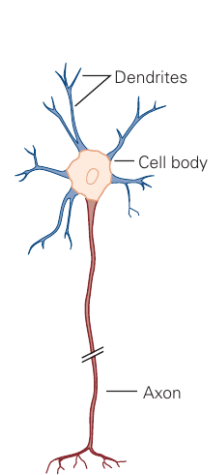
Bipolar cell of retina

C Pseudo-unipolar cell

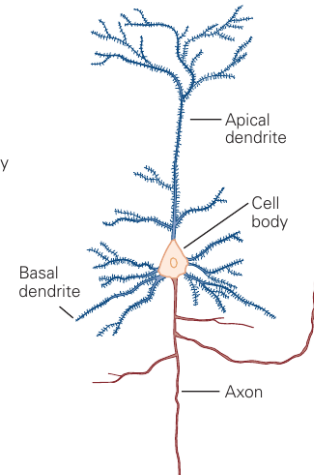


Ganglion cell of dorsal root

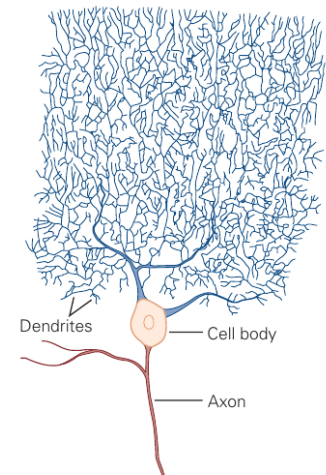
D Three types of multipolar cells



Motor neuron of spinal cord



Pyramidal cell of hippocampus



Purkinje cell of cerebellum

Neuronal shape varies with the inputs a cell receives and the connections it makes.

Kandel et al., Principles of Neural Science, 6th ed. (2021), Fig. 3-3, p. 60. Panels rearranged.

Further reading: foundations



Schultz W, Dayan P, Montague PR (1997). A Neural Substrate of Prediction and Reward. *Science* 275:1593–1599.

Sutton RS, Barto AG (2018). *Reinforcement Learning: An Introduction*, 2nd edition. MIT Press.

Steinberg EE et al. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience* 16:966–973.

Kobayashi S, Schultz W (2008). Influence of reward delays on responses of dopamine neurons. *Journal of Neuroscience* 28:7837–7846.

Bellemare MG, Dabney W, Munos R (2017). A Distributional Perspective on Reinforcement Learning. *ICML, PMLR* 70:449–458.

Dabney W et al. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature* 577:671–675.

Sutton RS, Barto AG (2018). Chapters 14–15: Psychology and Neuroscience. MIT Press.

Optional technical appendix

Further reading: recent papers



Mohebi A, Wei W et al. (2024). Dopamine transients follow a striatal gradient of reward time horizons. *Nature Neuroscience* 27:737–746.

Masset P et al. (2025). Multi-timescale reinforcement learning in the brain. *Nature* 642:682–690.

Lowet AS et al. (2025). An opponent striatal circuit for distributional reinforcement learning. *Nature* 639:717–726.

Sousa M et al. (2025). A multidimensional distributional map of future reward in dopamine neurons. *Nature* 642:691–699.

Lee RS et al. (2024). A feature-specific prediction error model explains dopaminergic heterogeneity. *Nature Neuroscience* 27:1574–1586.

Jang HJ et al. (2026). Acetylcholine demixes heterogeneous dopamine signals for learning and moving. *Nature Neuroscience* 29:840–850.

Optional technical appendix

Further reading: biology and control



Dayan P. Reinforcement Learning I: prediction and classical conditioning. Supplied lecture slides, rl_1.ppt. Concepts adapted; undated.

Dayan P. Reinforcement learning 2: action selection. Supplied lecture slides, rl_2.ppt. Concepts adapted and updated; undated.

Niv Y, Daw ND, Joel D, Dayan P (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 191:507–520.

Tritsch NX, Sabatini BL (2012). Dopaminergic modulation of synaptic transmission in cortex and striatum. *Neuron* 76:33–50.

Mohebi A et al. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature* 570:65–70.

Patriarchi T et al. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* 360:eaat4422.

Optional readings

Further reading: circuits and plasticity



Daw ND, Doya K (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology* 16:199–204.

Doya K (2008). Modulators of decision making. *Nature Neuroscience* 11:410–416.

Yagishita S et al. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345:1616–1620.

Collins AGE, Frank MJ (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review* 121:337–366.

Kandel ER, Koester JD, Mack SH, Siegelbaum SA, eds. (2021). *Principles of Neural Science*, 6th ed. McGraw Hill.

Picciotto MR, Higley MJ, Mineur YS (2012). Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron* 76:116–129.

Matsuda W et al. (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *Journal of Neuroscience* 29:444–453.

Selected sources