COMPSCI 389
Introduction to Machine Learning

Days: Tu/Th. **Time:** 2:30 – 3:45 **Building:** Morrill 2 **Room:** 222

Topic 14.0: Psychology and Neuroscience

Prof. Philip S. Thomas (pthomas@cs.umass.edu)

Machine learning has had a profound impact on other fields.

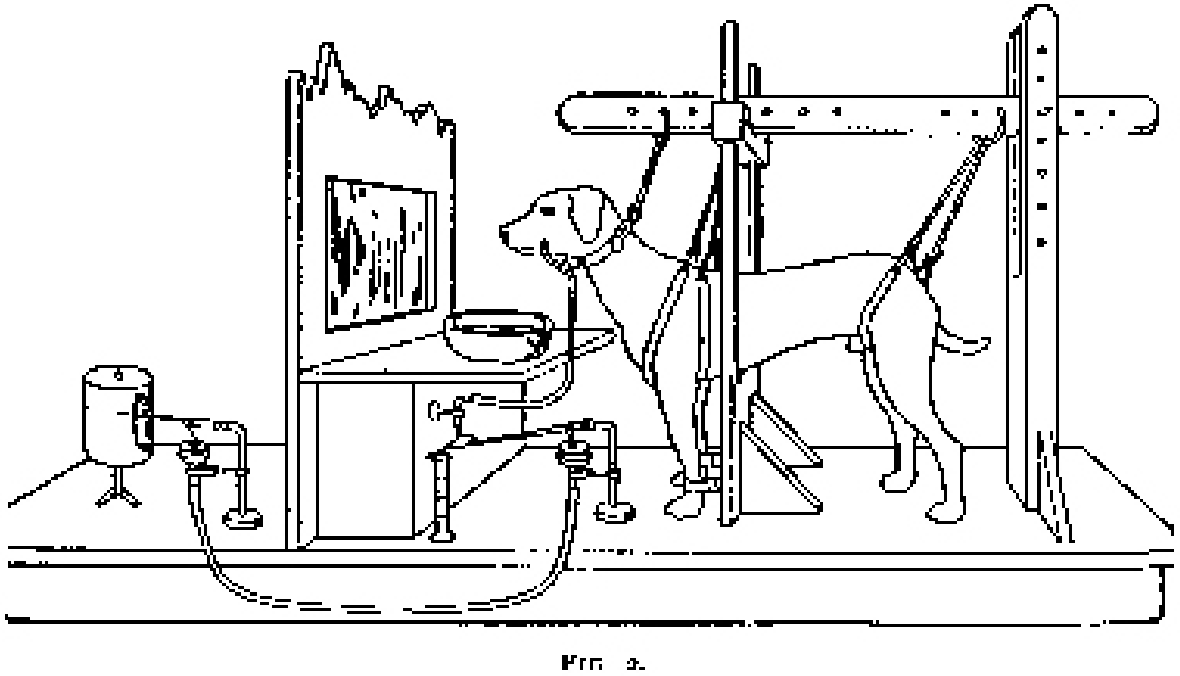
Other fields have had a profound impact on machine learning.

(Behavioral) Psychology

- **Operant Conditioning**: Learning process through which the strength of a behavior is modified by reward or punishment.
  - "**Control**": Learning a policy that maximizes the expected sum of rewards
  - In psychology, operant conditioning is sometimes called *instrumental conditioning*.
- **Classical Conditioning**: Learning procedure in which a biologically potent stimulus (e.g., food) is paired with a previously neural stimulus (e.g., a bell).
  - *Not* about learning how to get more reward
  - "**Prediction/Evaluation**": Learning a value function

Classical Conditioning (Example)

- In 1897, Ivan Pavlov found that dogs began to salivate when they observe cues that indicate they will be fed
  - This is before they are shown food (**unconditioned stimulus**).
- Pavlov played a sound before feeding a dog.
- The dog began to salivate in response to this sound (**conditioned stimulus**).



Operant Conditioning (Example)

- In 1898 Edward Thorndike observed the behavior of cats in “puzzle boxes.”

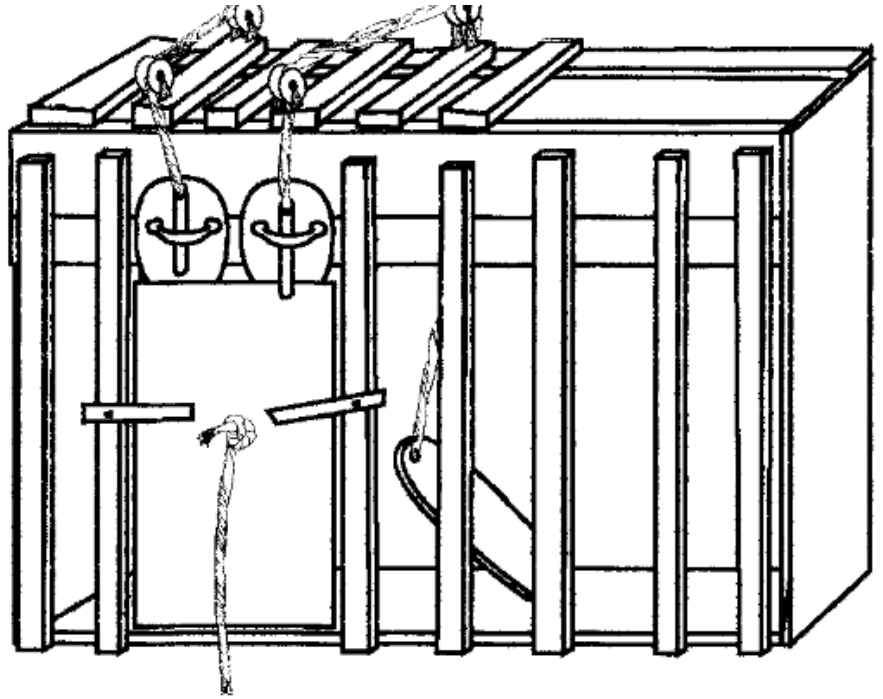


Fig. 4. Box K. The door is held in place by a weight suspended by a string. To open the door, a cat had to depress a treadle, pull on a string, and push a bar up or down. (After Thorndike, 1898, Figure 1, p. 8.)

Figure from “Thorndike’s Puzzle Boxes and the Origins of The Experimental Analysis of Behavior” by Paul Chance, 1999.

Operant Conditioning (Example)

- In 1898 Edward Thorndike observed the behavior of cats in “puzzle boxes.”
  - The cats had to follow a sequence of actions to escape from the box.
  - Thorndike timed how long it took the cats to escape the puzzle box each time they were placed in it.
- Thorndike found that the time it took cats to escape tended to decrease from 300 seconds to ~6-7 seconds.

> The cat that is clawing all over the box in her impulsive struggle will probably claw the string or loop or button so as to open the door. And gradually all the other non-successful impulses will be stamped out and the particular impulse leading to the successful act will be stamped in by the resulting pleasure, until, after many trials, the cat will, when put in the box, immediately claw the button or loop in a definite way. (Thorndike 1898, p. 13)

Excerpt from Sutton & Barto, second edition, quoting Thorndike.

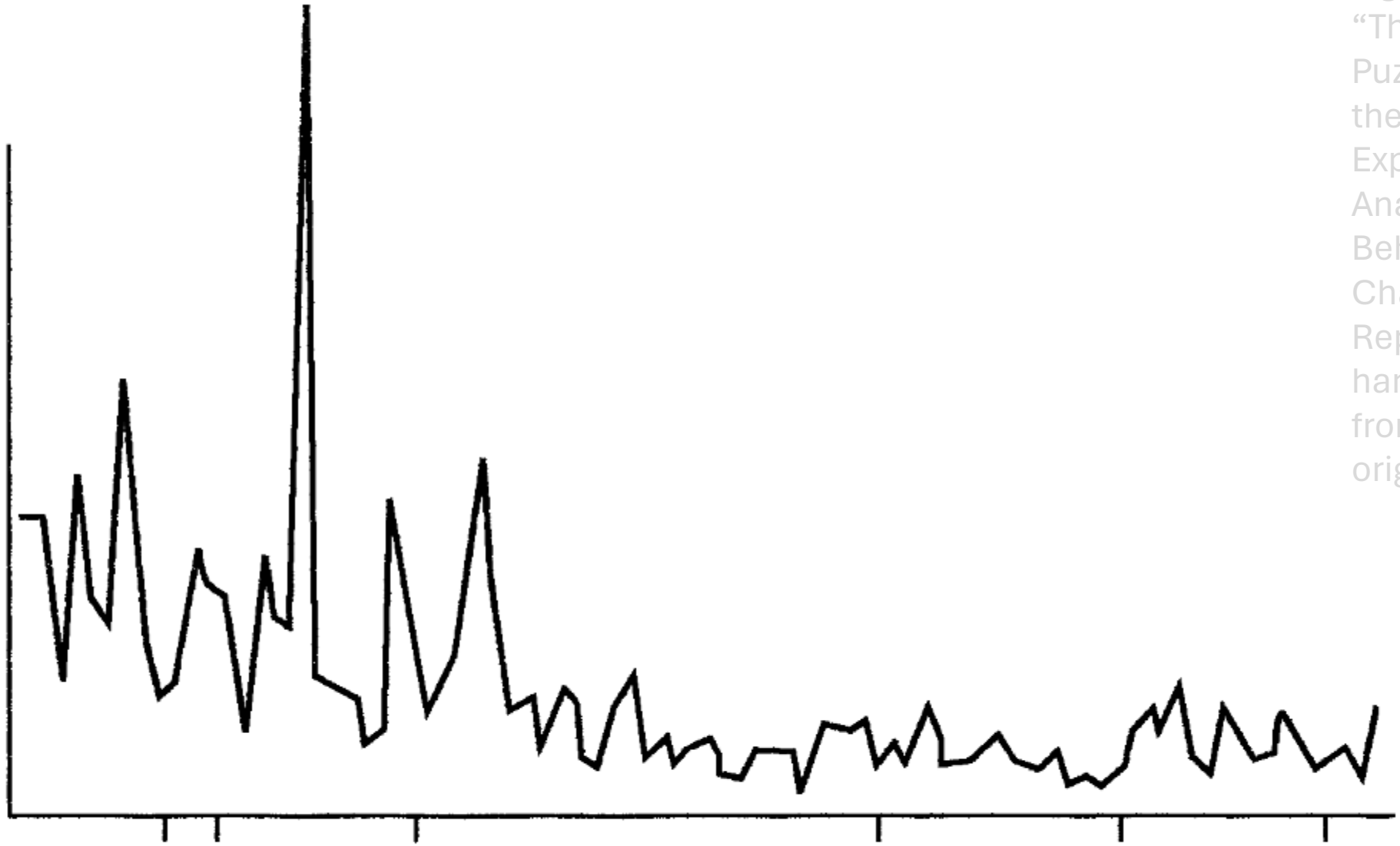


Fig. 6. Performance of Cat 4 in Box K. Box K required three distinct responses. The figure shows escape times on approximately 117 trials over a 7-day period. Progress was slow and erratic. (After Thorndike, 1898, Figure 10, p. 26.)

Figure from "Thorndike's Puzzle Boxes and the Origins of The Experimental Analysis of Behavior" by Paul Chance, 1999. Reproduction of a hand-drawn figure from Thorndike's original work.

Behaviorism

- John B. Watson is credited with founding behaviorism in a 1913 paper “Psychology as the Behaviorist Views It”.
- B. F. Skinner later expanded on Watson’s ideas with his own experiments on operant conditioning similar to Thorndike’s.
- Skinner coined the phrases operant conditioning and classical conditioning.
- Andy Barto made it clear that RL was inspired by this work in behaviorist psychology. Recall our definition of RL:

Reinforcement learning is an area of machine learning, inspired by behaviorist psychology, concerned with how an agent can learn from interactions with an environment

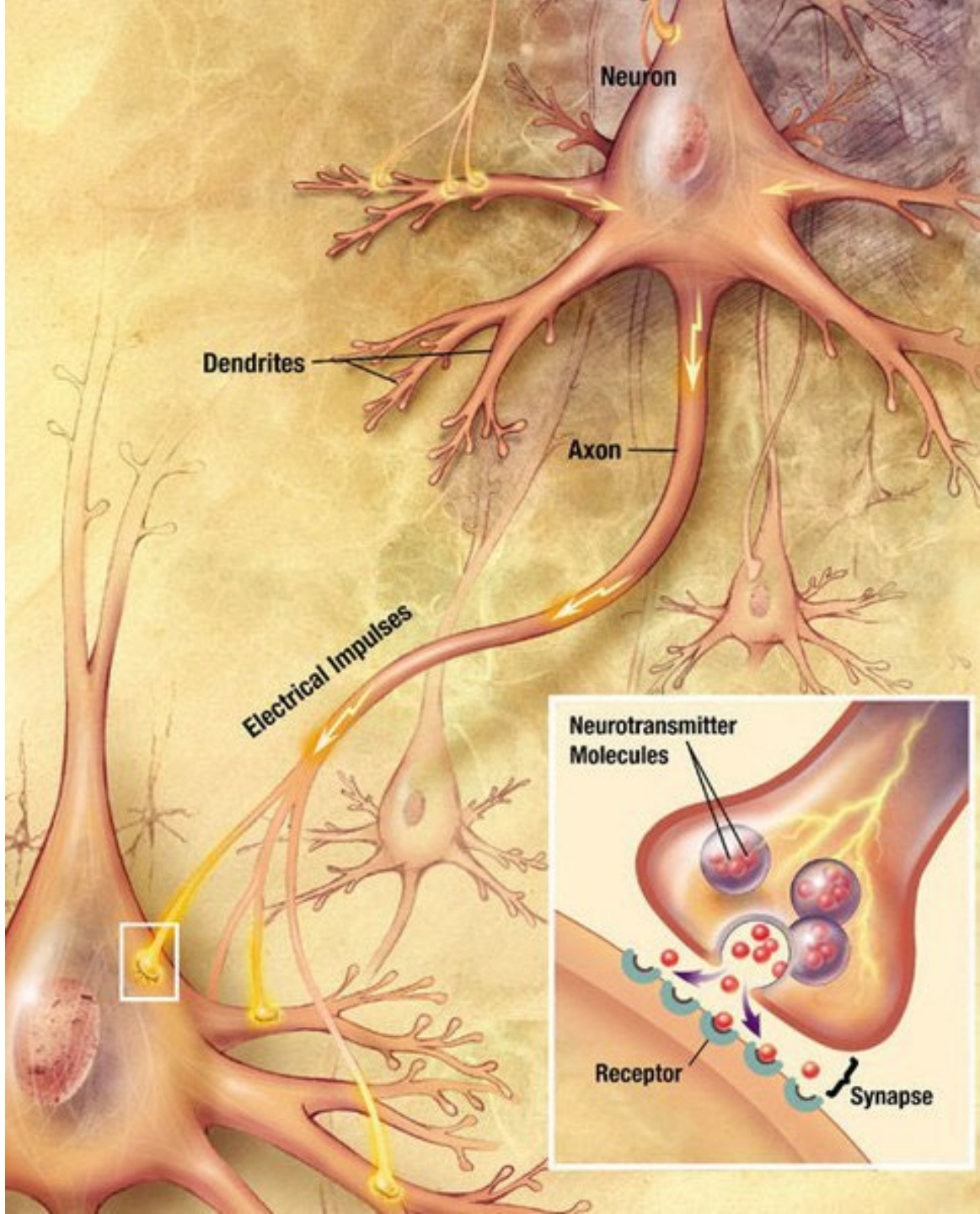
– Wikipedia / Sutton&Barto / Phil

Neuroscience

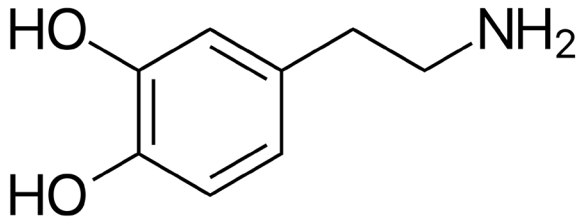
- Animals and RL agents solve similar problems: learning, via trial-and-error, how to maximize the amount of reward they receive.
- Neuroscience studies how this learning occurs in animals.
- RL studies the mathematical and algorithmic foundations of this type of learning.
- It would be surprising if there were *not* similarities between the algorithms that have been found to be effective in RL and the learning mechanisms created by evolution.

Neurotransmitters

- A neurotransmitter is a chemical that is transmitted between neurons.
- The **presynaptic neuron** is the neuron that releases the neurotransmitter.
- The **postsynaptic neuron** is the neuron that receives the neurotransmitter.



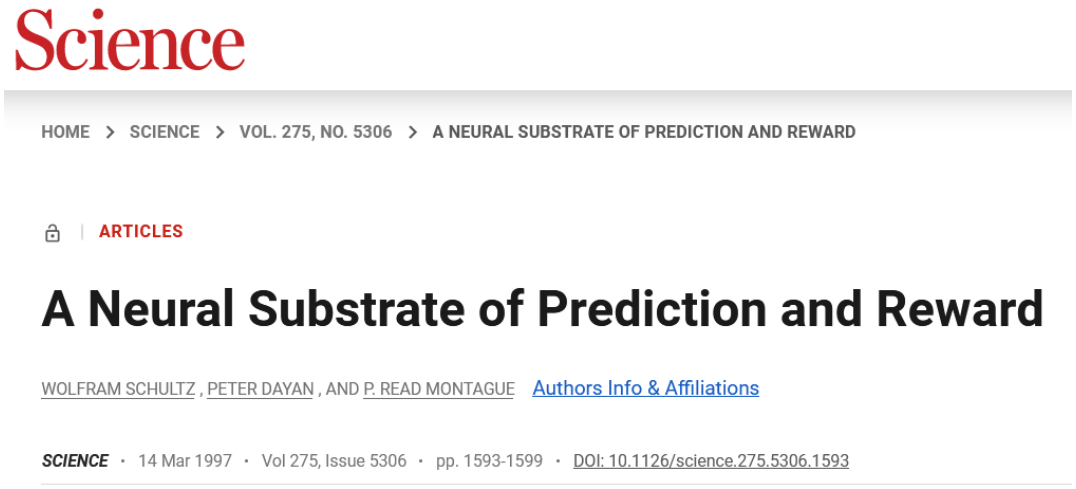
Dopamine



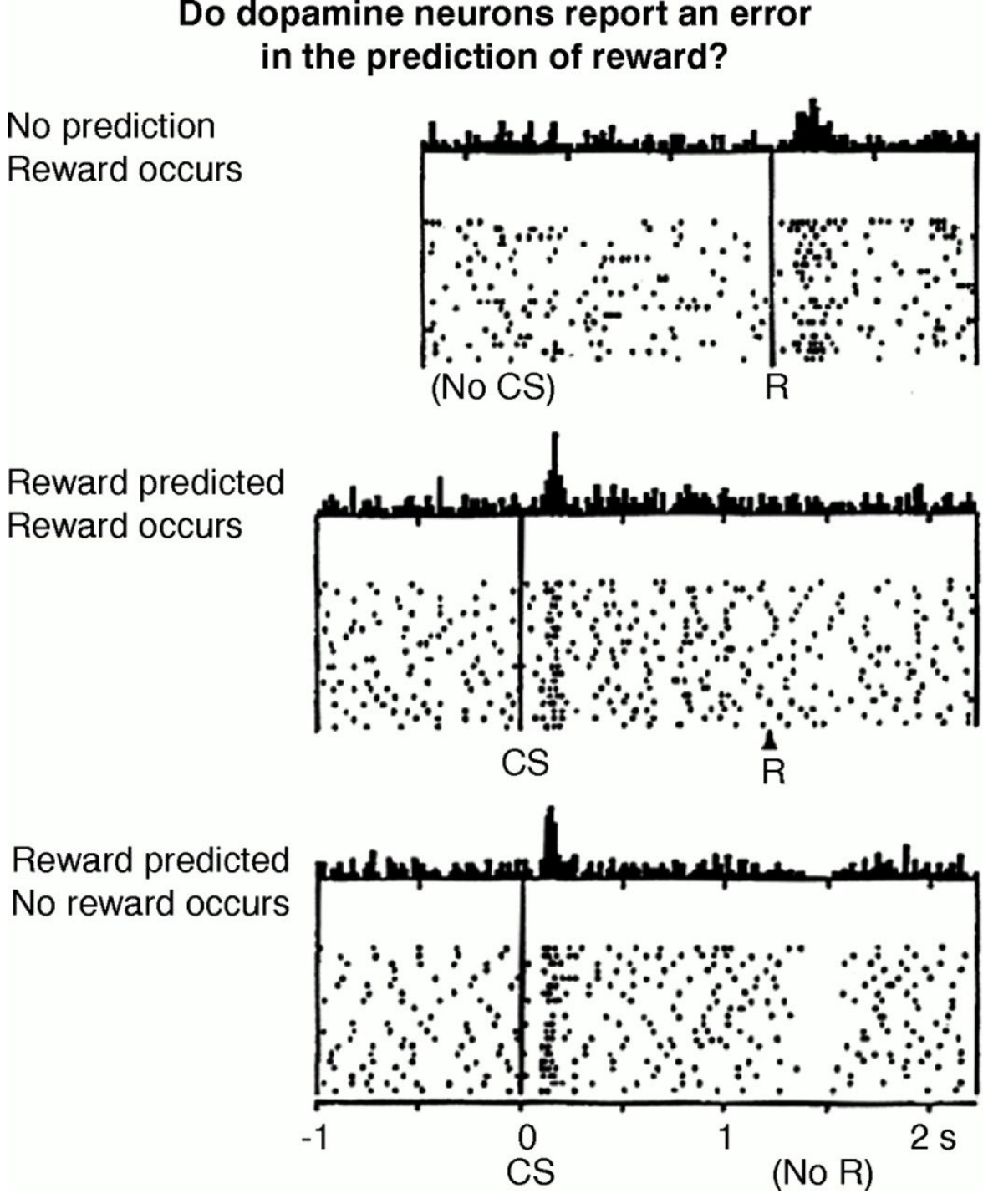
- There are at least 100 different neurotransmitters.
- Different neurotransmitters can play different roles in the brain (and the body).
- **Dopamine** (3,4-dihydroxyphenethylamine) is one such neurotransmitter**.**
- By the mid 1900s, neuroscientists knew that dopamine plays a role in learning and relates to rewards.
  - E.g., Olds & Miller hypothesized that dopamine ≈ reward in 1954.
  - **Note**: Dopamine can be found in several parts of your body, where it has different roles. For example, in your kidneys it helps to regulate blood flow and sodium excretion. This discussion is only about dopamine as a neurotransmitter.

The role of dopamine

- Discussions in the 1980s and 1990s between reinforcement learning researchers and neuroscientists resulted in hypotheses that *dopamine* may be the neural correlate of *temporal difference error*.
- The first strong evidence for this hypothesis was published by Shultz, Dayan, and Montague in *Science*



- **Key insight**: dopamine doesn't encode rewards, it encodes **reward prediction errors** (RPEs).
- This is called the *reward prediction error hypothesis for dopamine*.
- The paper discusses how this relates to temporal difference error in RL.

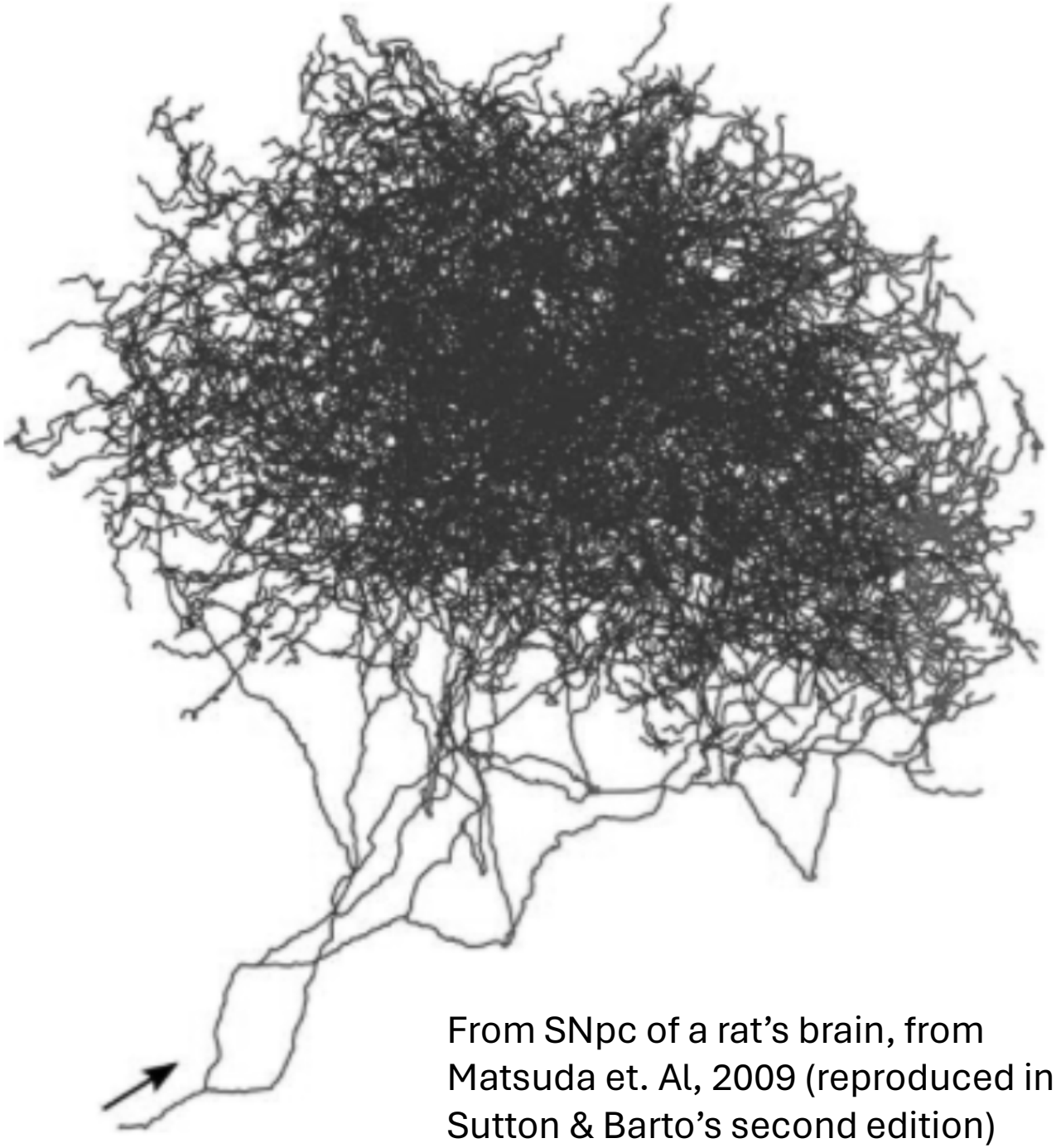


Dopaminergic Neurons

- Neurons that excrete the neurotransmitter dopamine are called **dopaminergic neurons**.
- In humans, these neurons can be found in:
  - Substantia Nigra pars Compacta (SNpc)
    - Sends dopamine to parts of the *striatum*, which relates to motor and action planning, decision making, and motivation.
  - Ventral Tegmental Area (VTA)
    - Sends dopamine to the prefrontal cortex, which relates to planning, personality, and decision making.

Axonal arbor of a dopaminergic neuron

- Dopaminergic neurons have particularly large **axonal arbors** (sets of connections to other neurons).
  - 100 to 1000 times more than typical neurons.
  - *Each* dopaminergic neuron makes roughly 500,000 connections.
- Dopamine appears to be “broadcast” to large portions of the brain, like TD-error in an actor-critic.



From SNpc of a rat’s brain, from Matsuda et. Al, 2009 (reproduced in Sutton & Barto’s second edition)

RPE hypothesis for dopamine

- The RPE hypothesis remains a hypothesis.
- Some studies suggest that the relationship only holds for positive TD errors [1,4].
  - Others suggest that different dopaminergic neurons may encode positive and negative TD-errors [5]
- Researchers have questioned whether dopamine has a *causal* impact on behavior, but there is mounting evidence that it does [6].
- Some research suggests that dopamine corresponds to variants of TD-error, like those from *distributional RL* [3].
- It is unclear how the relationship between dopamine and TD-error extends across the animal kingdom.
  - There is mounting evidence that dopamine plays similar roles in mammals [7] and even flies (albeit with the sign reversed) [2].

[1] H. M. Bayer and P. W. Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–141, 2005.

[2] A. Claridge-Chang, R. D. Roorda, E. Vrontou, L. Sjulson, H. Li, J. Hirsh, and G. Miesenböck. Writing memories with light-addressable reinforcement circuitry. *Cell*, 139(2):405–415, 2009.

[3] W. Dabney, Z. Kurth-Nelson, N. Uchida, C. K. Starkweather, D. Hassabis, R. Munos, and M. Botvinick. A distributional code for value in dopamine-based reinforcement learning. *Nature*, 577(7792):671–675, 2020.

[4] K. D'Ardenne, S. M. McClure, L. E. Nystrom, and J. D. Cohen. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319(5867):1264–1267, 2008.

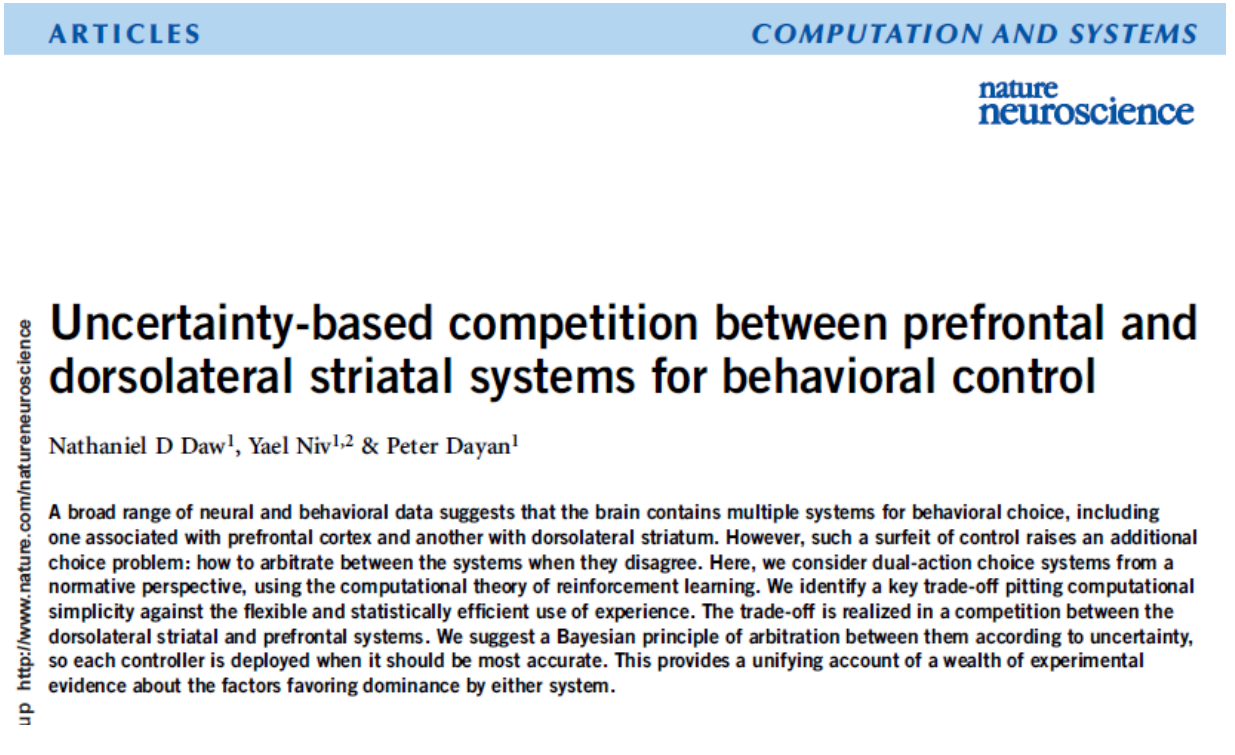
[5] M. Matsumoto and O. Hikosaka. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, 459(7248):837–841, 2009.

[6] E. E. Steinberg, R. Keiflin, J. R. Boivin, I. B. Witten, K. Deisseroth, and P. H. Janak. A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience*, 16(7):966–973, 2013.

[7] S. Waddell. Reinforcement signalling in Drosophila; dopamine does it all after all. *Current Opinion in Neurobiology*, 23(3):324–329, 2013.

RL & Neuroscience: Other Topics

- There is a wide range of other research at the intersection of RL and neuroscience.
- One example: reward devaluation studies.



ARTICLES COMPUTATION AND SYSTEMS

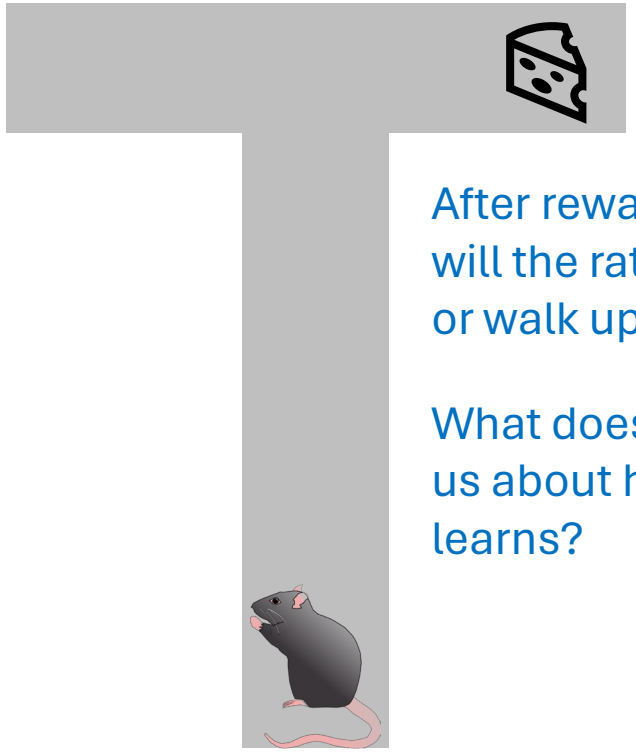
nature neuroscience

Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control

Nathaniel D Daw[1], Yael Niv[1,2] & Peter Dayan[1]

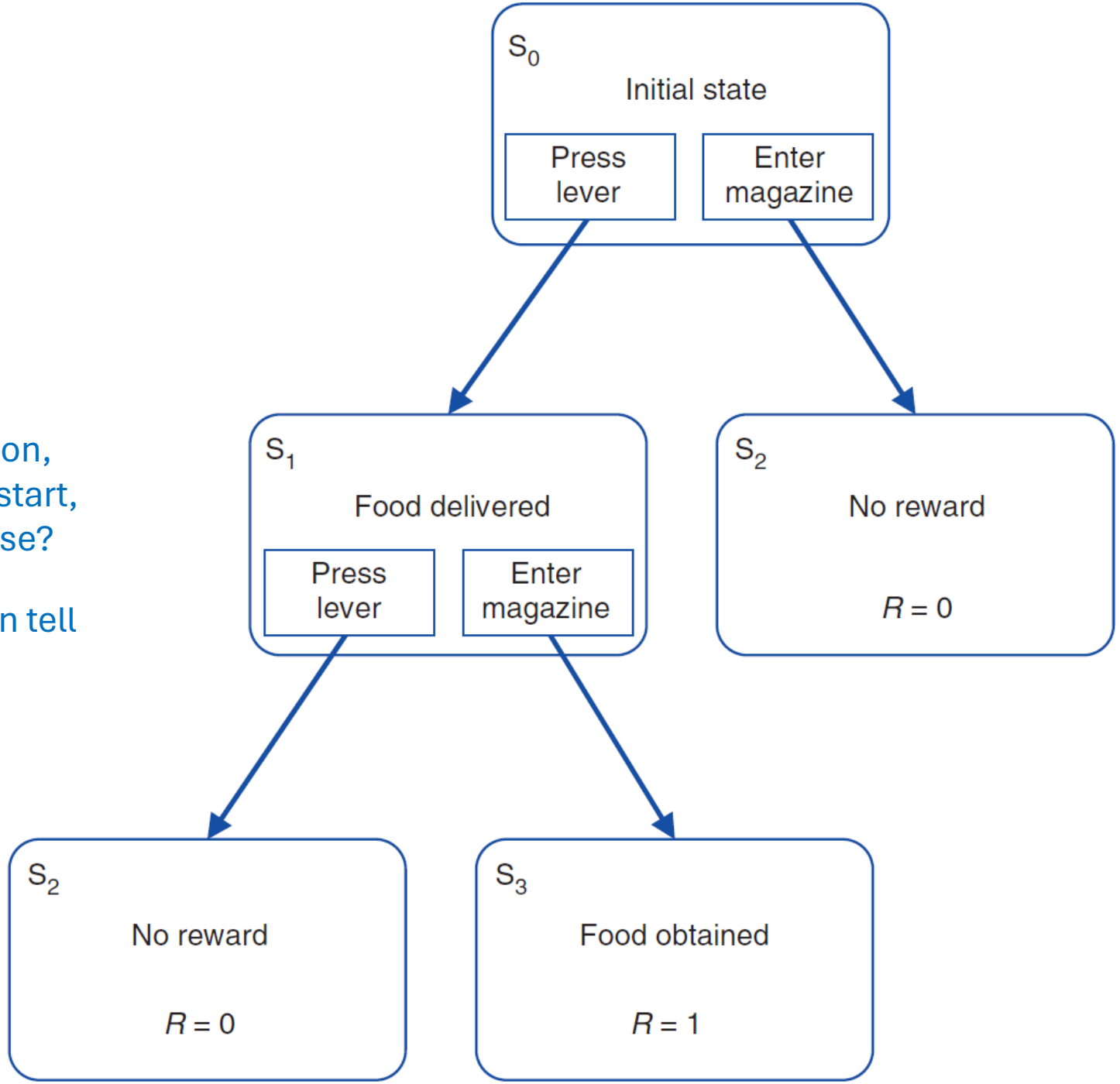
A broad range of neural and behavioral data suggests that the brain contains multiple systems for behavioral choice, including one associated with prefrontal cortex and another with dorsolateral striatum. However, such a surfeit of control raises an additional choice problem: how to arbitrate between the systems when they disagree. Here, we consider dual-action choice systems from a normative perspective, using the computational theory of reinforcement learning. We identify a key trade-off pitting computational simplicity against the flexible and statistically efficient use of experience. The trade-off is realized in a competition between the dorsolateral striatal and prefrontal systems. We suggest a Bayesian principle of arbitration between them according to uncertainty, so each controller is deployed when it should be most accurate. This provides a unifying account of a wealth of experimental evidence about the factors favoring dominance by either system.

up http://www.nature.com/natureneuroscience

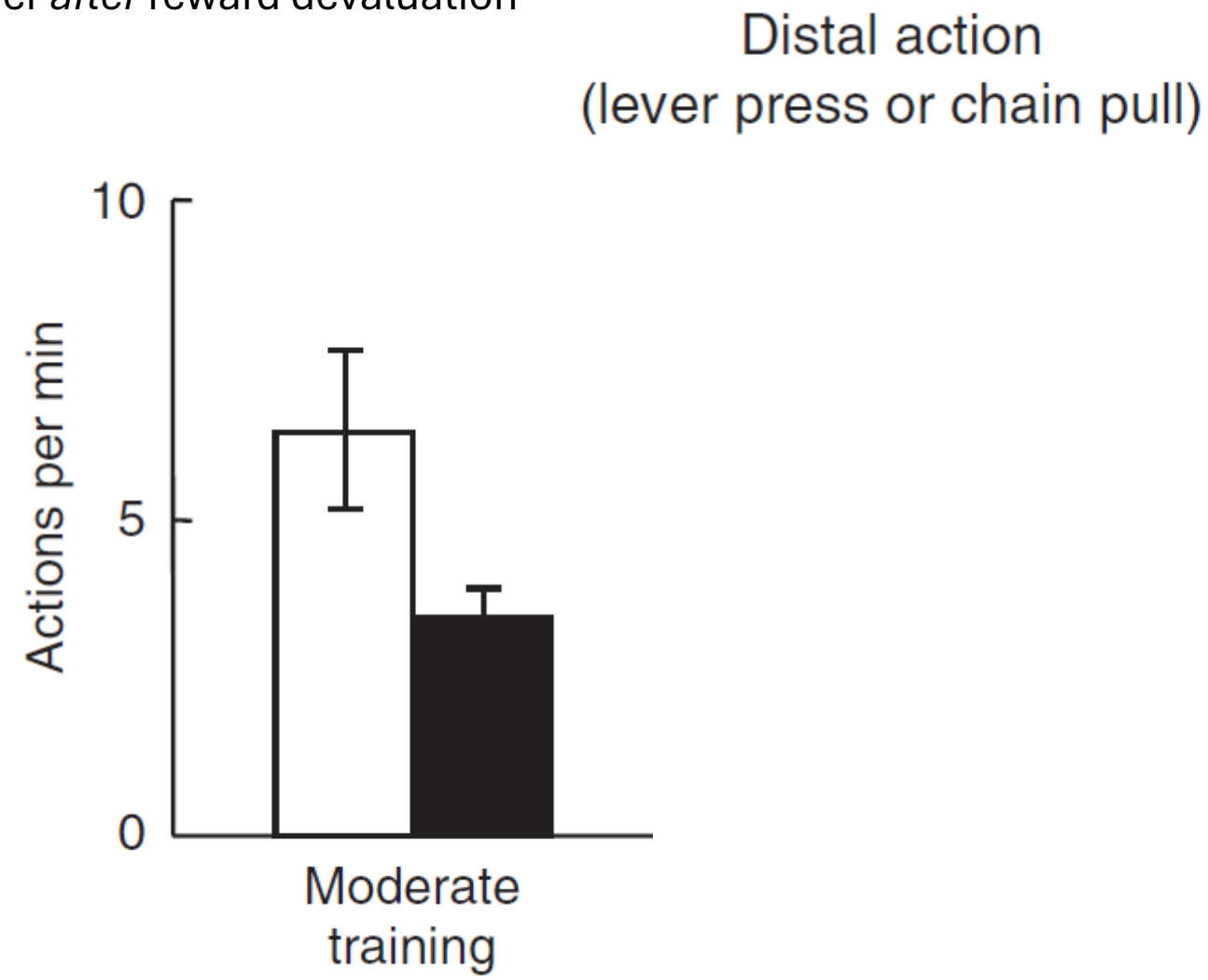


After reward devaluation, will the rat stay at the start, or walk up to the cheese?

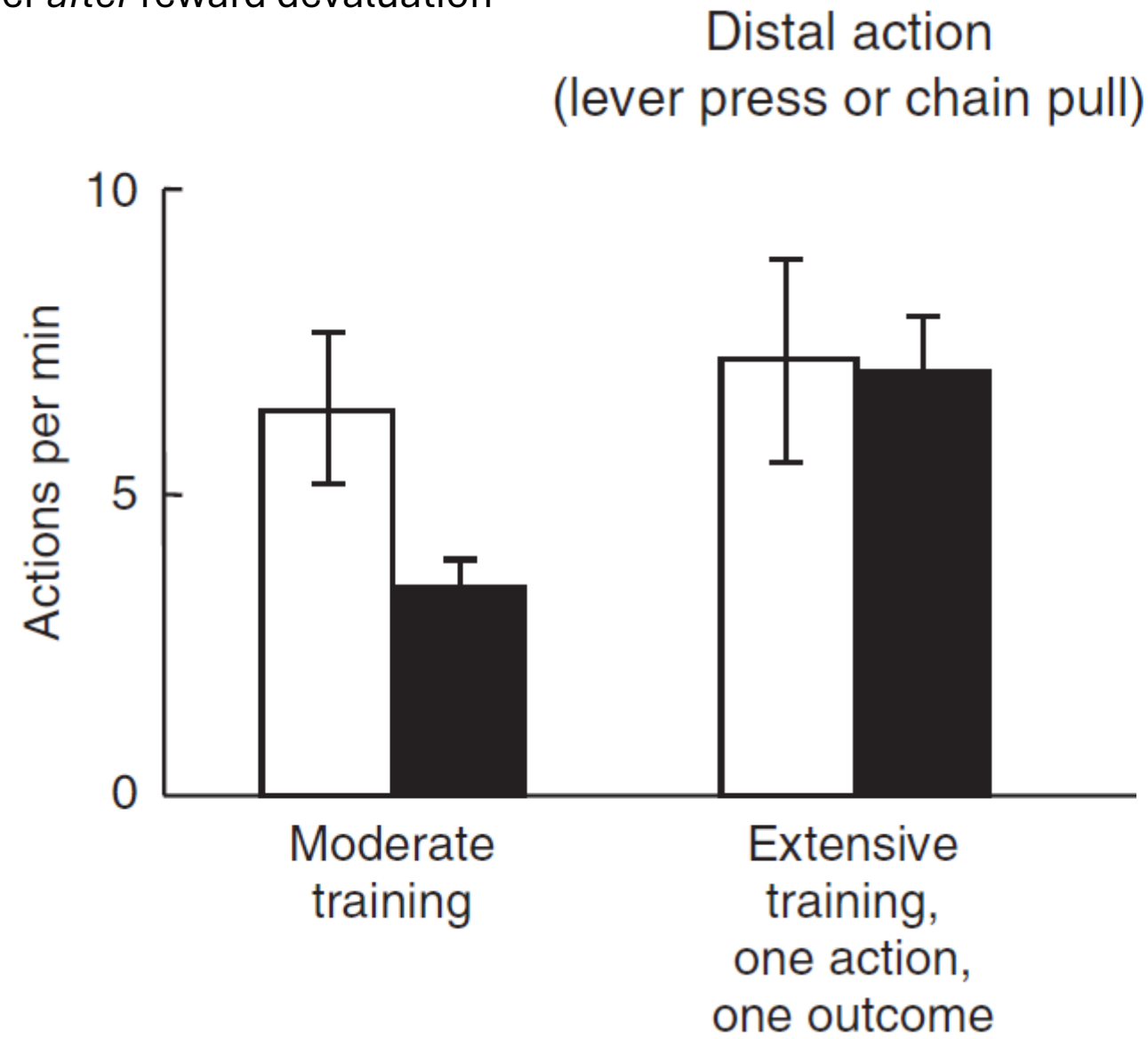
What does each option tell us about how the rat learns?



White = hit the lever *before* reward devaluation
Black = hit the lever *after* reward devaluation



White = hit the lever *before* reward devaluation
Black = hit the lever *after* reward devaluation

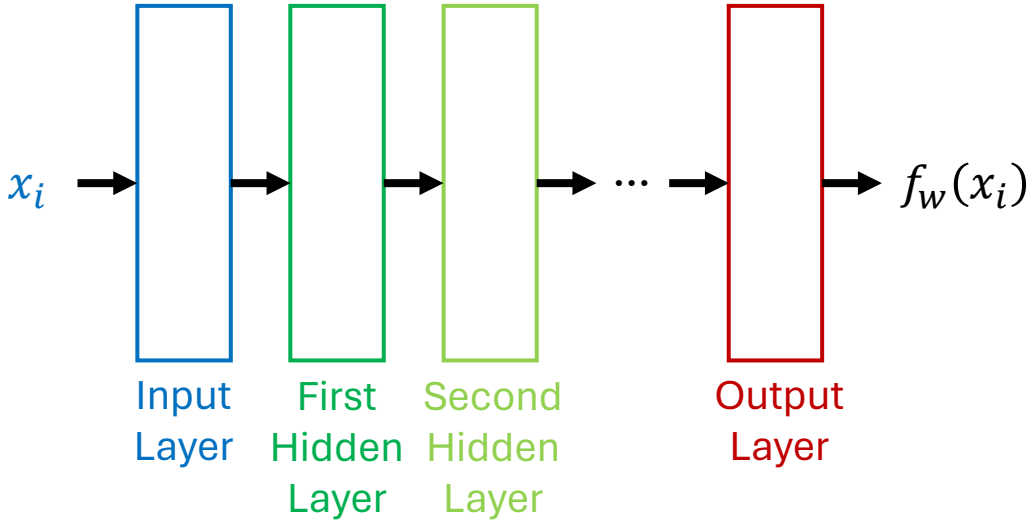


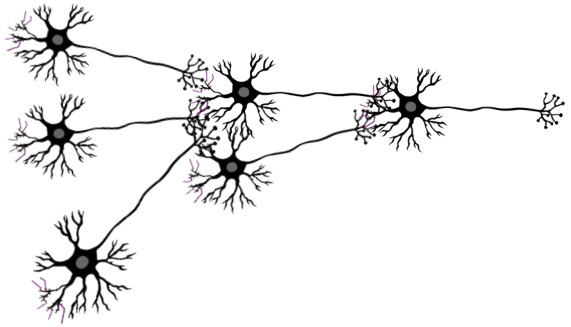
Note: More details and variants studied in the paper (e.g., multiple actions/outcomes).

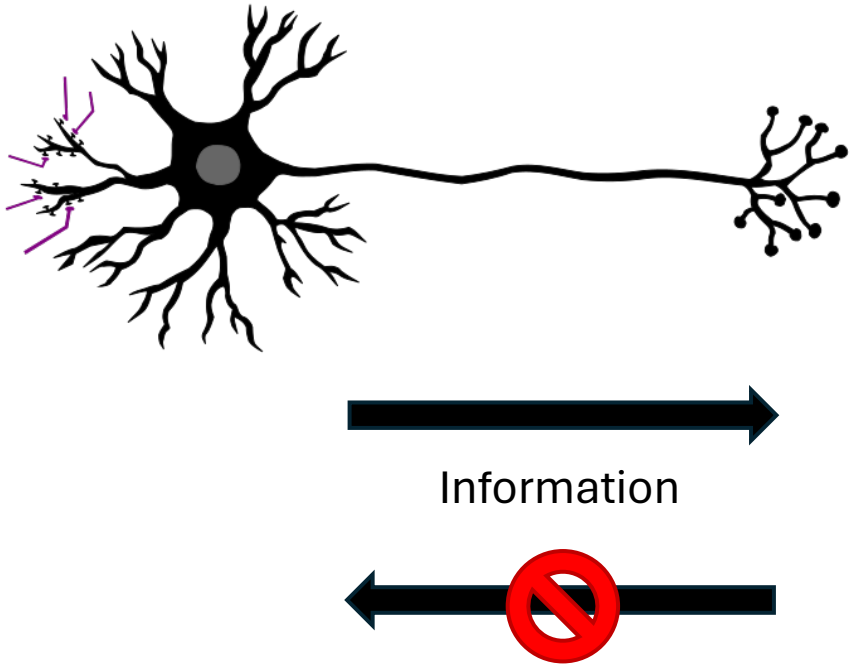
Conclusion: Early learning appears to be model-based planning, which transitions to a model-free policy over time.

RL & Neuroscience: Other Topics

- Brains (probably) do not implement backpropagation!



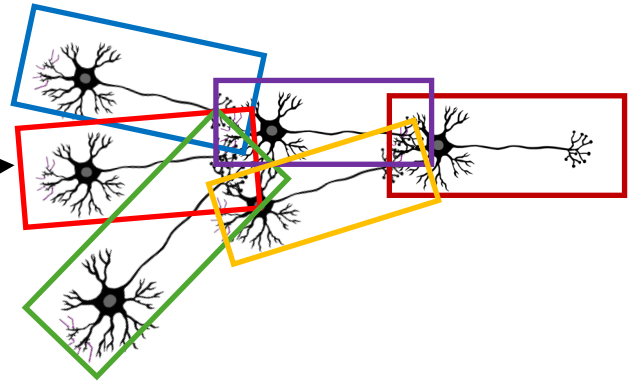




Information does not flow backwards down the axon!

RL & Neuroscience: Other Topics

- Brains (probably) do not implement backpropagation!
  - Information does not flow backwards down the axon.
  - Some computer scientists have tried to argue for alternate implementations of backpropagation (e.g., each neuron has a corresponding neuron that has the reverse connections and passes information backwards through the network).
    - These hypotheses have not been well-received by the neuroscience community.
- Some ML research tries to study more biologically plausible ways of training parametric models.
  - Spiking networks
  - Coagent networks



Each neuron (or group of neurons) can be viewed as an independent RL agent!

RL & Neuroscience: Other Topics

- Reinforcement learning has been used to model addiction (sometimes via the $\delta_t \leftrightarrow$ dopamine connection).
- There is an entire conference, *Reinforcement Learning and Decision Making* (RLDM) devoted to bringing RL researchers and neuroscientists together to learn from each other!

End