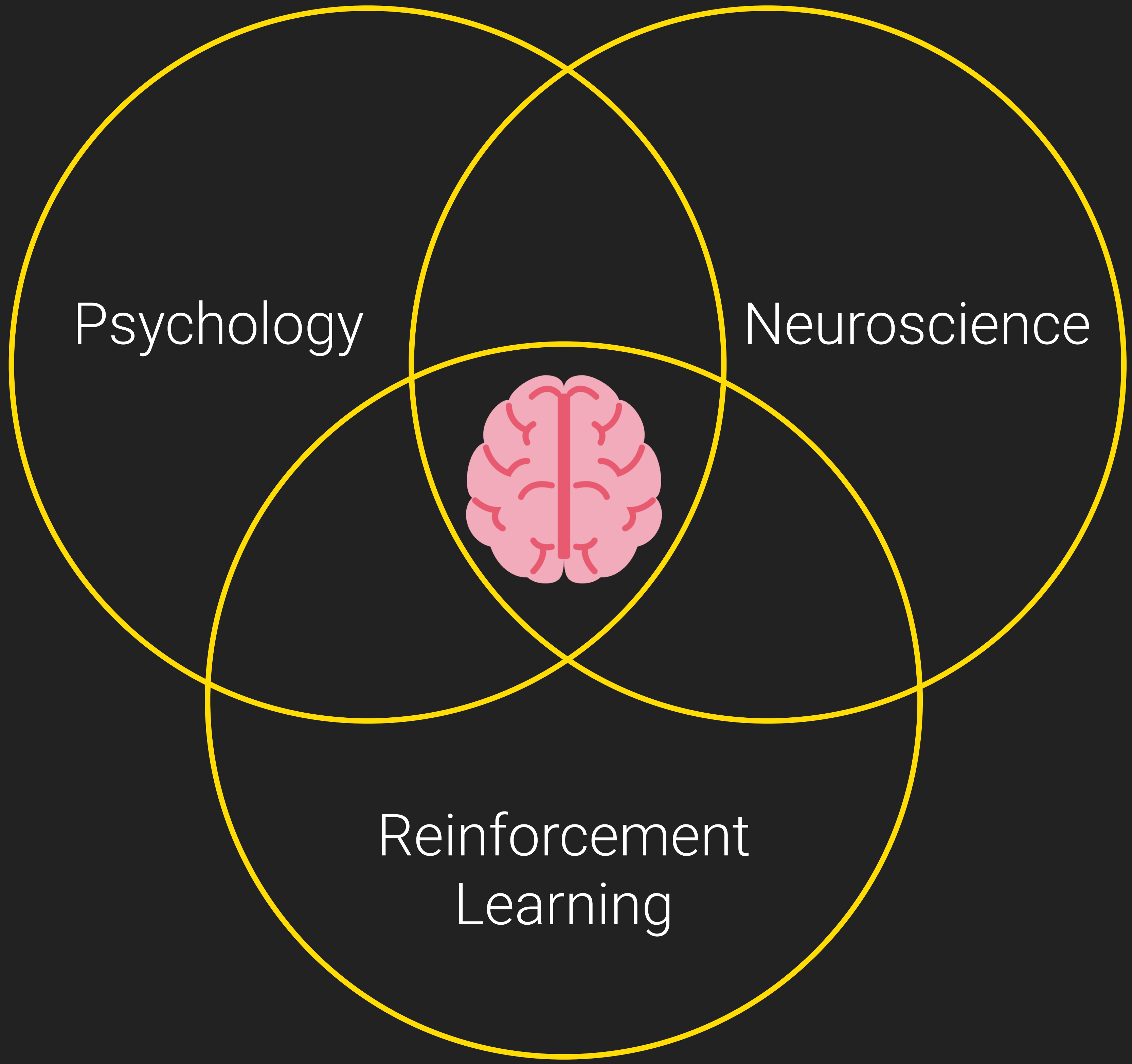
FIGURING OUT HOW THE MIND WORKS

At the Exciting Intersection of RL, Psychology, and Neuroscience

Abhishek Naik



WE ARE ALL STUDYING HOW THE BRAIN WORKS

OUTLINE

OUTLINE

- Some RL and Psychology background
    - The RL framework, TD learning, Blocking, Higher-order Conditioning

OUTLINE

- Some RL and Psychology background
    - The RL framework, TD learning, Blocking, Higher-order Conditioning
- Connections between RL and Psychology
    - TD model explains and predicts psychological phenomena

OUTLINE

- Some RL and Psychology background
  - The RL framework, TD learning, Blocking, Higher-order Conditioning
- Connections between RL and Psychology
  - TD model explains and predicts psychological phenomena
- Some Neuroscience background
  - Action potentials, Dopamine

OUTLINE

- Some RL and Psychology background
    - The RL framework, TD learning, Blocking, Higher-order Conditioning
- Connections between RL and Psychology
    - TD model explains and predicts psychological phenomena
- Some Neuroscience background
    - Action potentials, Dopamine
- Connections between RL and Neuroscience
    - The Reward Prediction Error Hypothesis

OUTLINE

- Some RL and Psychology background
  - The RL framework, TD learning, Blocking, Higher-order Conditioning
- Connections between RL and Psychology
  - TD model explains and predicts psychological phenomena
- Some Neuroscience background
  - Action potentials, Dopamine
- Connections between RL and Neuroscience
  - The Reward Prediction Error Hypothesis
- Where do we go from here?
  - Opportunities for collaborations among all these disciplines

SUPERVISED LEARNING VS REINFORCEMENT LEARNING

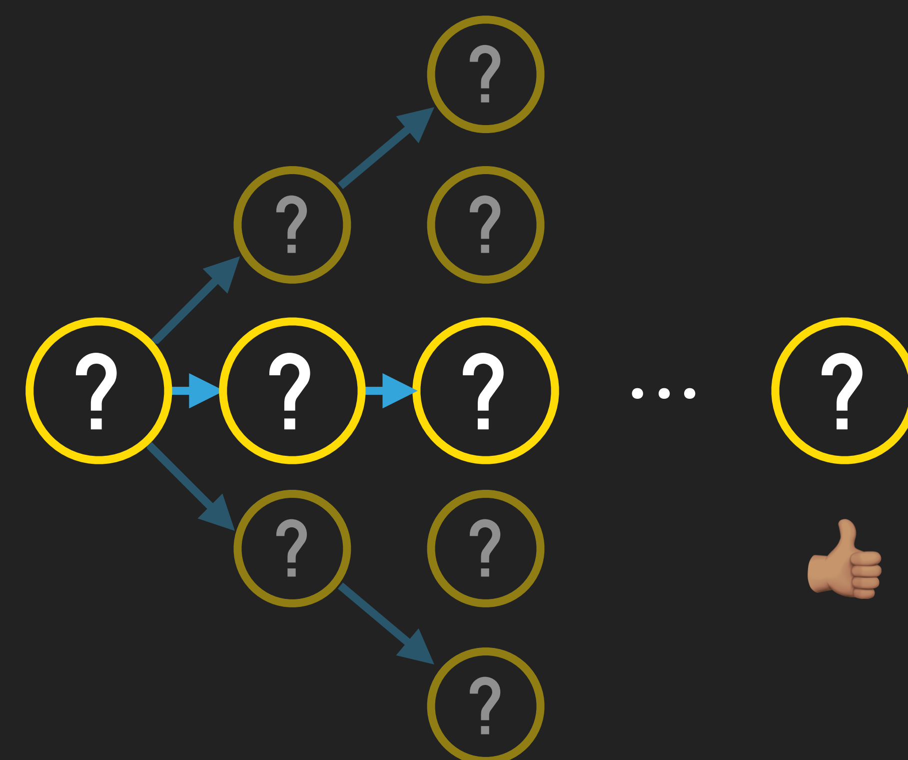
SUPERVISED LEARNING VS REINFORCEMENT LEARNING

- Independent decisions
- Instructive feedback
- Immediate feedback

SUPERVISED LEARNING VS REINFORCEMENT LEARNING

- Independent decisions
- Instructive feedback
- Immediate feedback

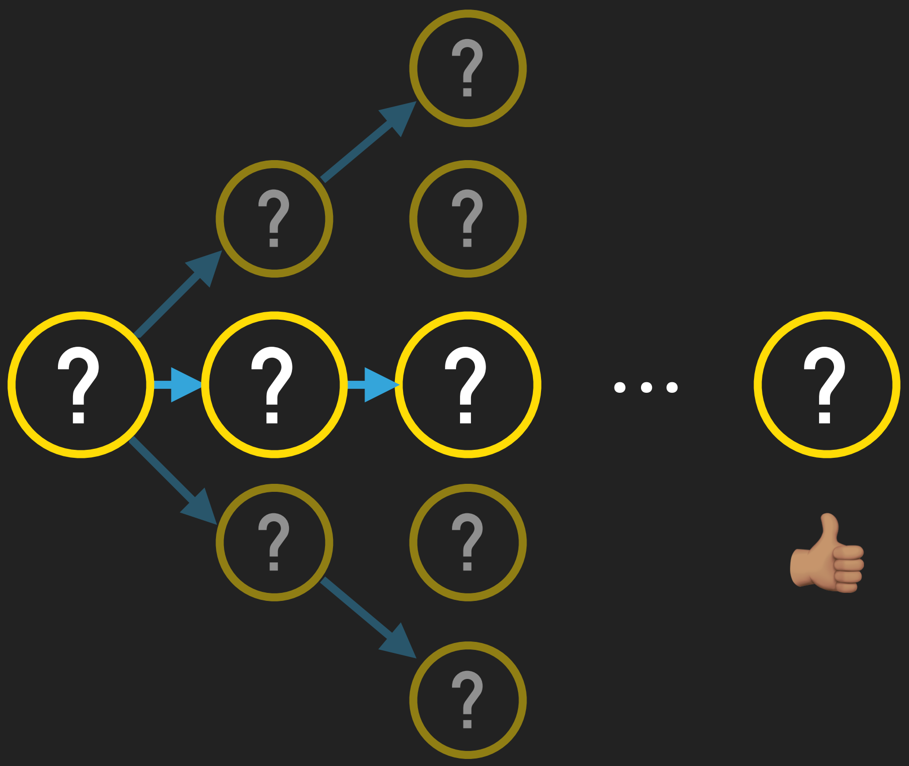
- Sequential decisions
- Evaluative feedback
- Delayed feedback



SUPERVISED LEARNING VS REINFORCEMENT LEARNING

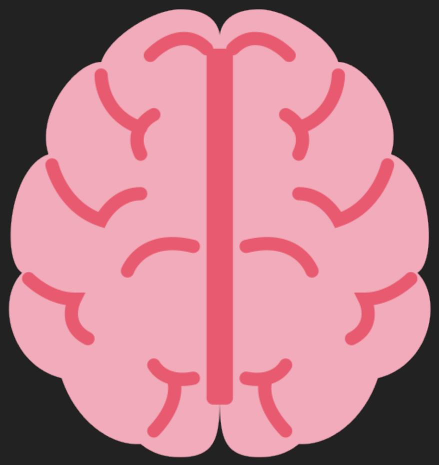
- Independent decisions
- Instructive feedback
- Immediate feedback

- Sequential decisions
- Evaluative feedback
- Delayed feedback

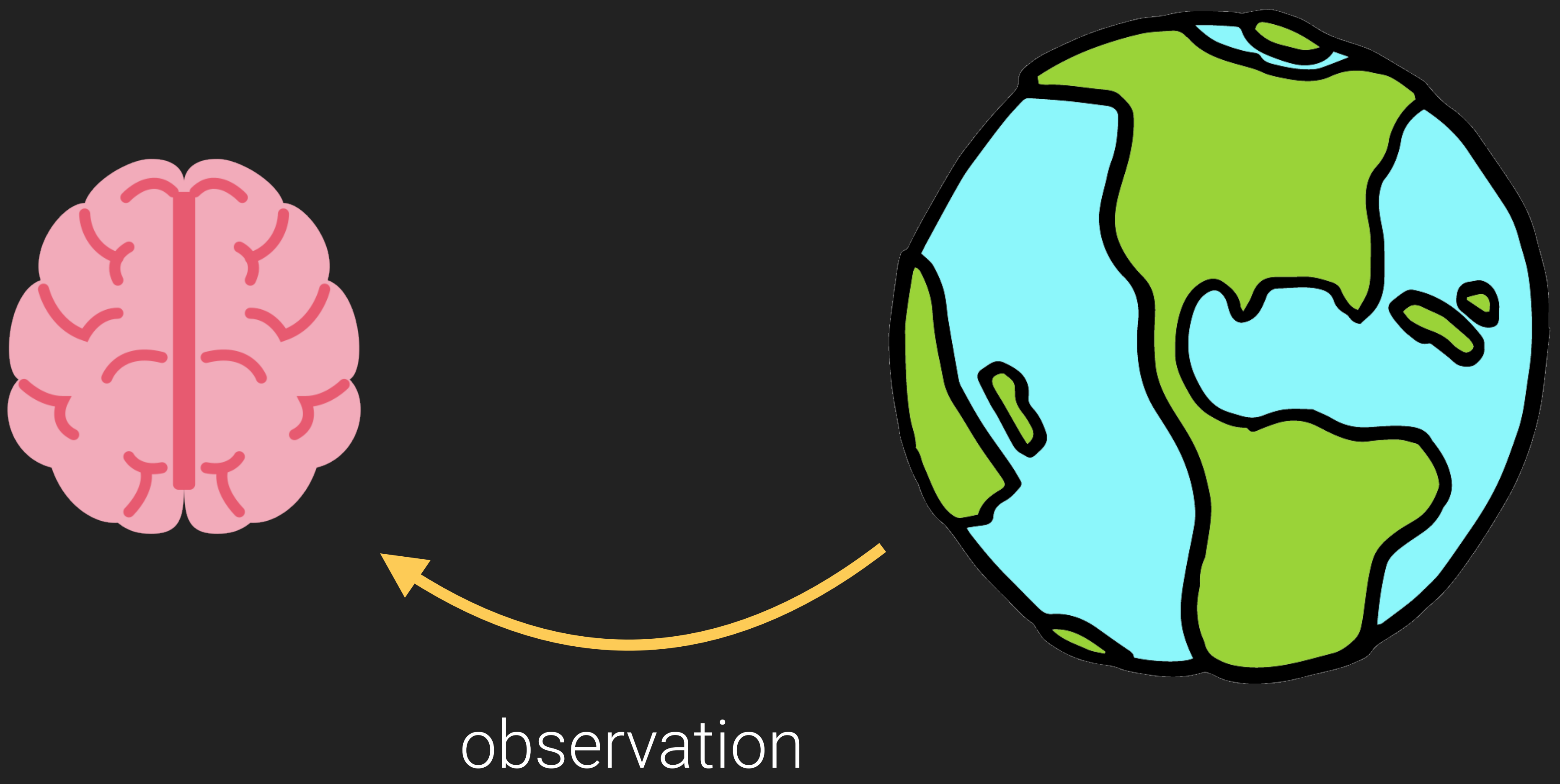


MOST REAL-WORLD TASKS ARE SEQUENTIAL IN NATURE

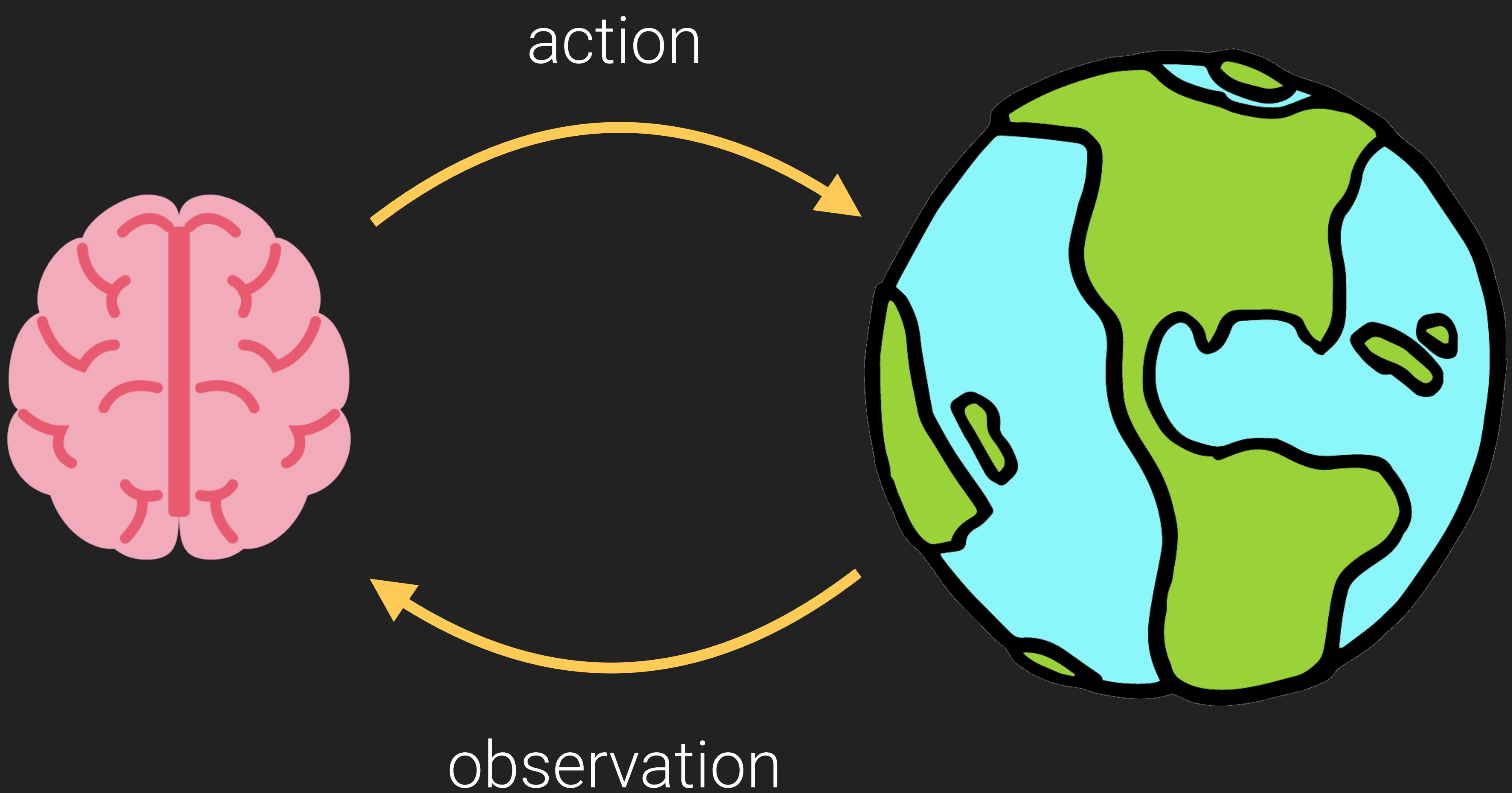
THE RL FRAMEWORK



THE RL FRAMEWORK



THE RL FRAMEWORK



TEMPORAL-DIFFERENCE (TD) LEARNING

TEMPORAL-DIFFERENCE (TD) LEARNING

- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

TEMPORAL-DIFFERENCE (TD) LEARNING

- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{correct})$$

TEMPORAL-DIFFERENCE (TD) LEARNING

- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{correct})$$

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{better})$$

TEMPORAL-DIFFERENCE (TD) LEARNING

- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

$$P_{new} = (1-\alpha)P_{old} + \alpha(P_{correct})$$

$$P_{new} = (1-\alpha)P_{old} + \alpha(P_{better})$$

$$= P_{old} + \alpha(P_{better} - P_{old})$$

TEMPORAL-DIFFERENCE (TD) LEARNING

- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{correct})$$

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{better})$$

$$= P_{old} + \alpha(P_{better} - P_{old})$$

$$V_{new}(s) = V_{old}(s) + \alpha(R + V_{old}(s') - V_{old}(s))$$

TEMPORAL-DIFFERENCE (TD) LEARNING

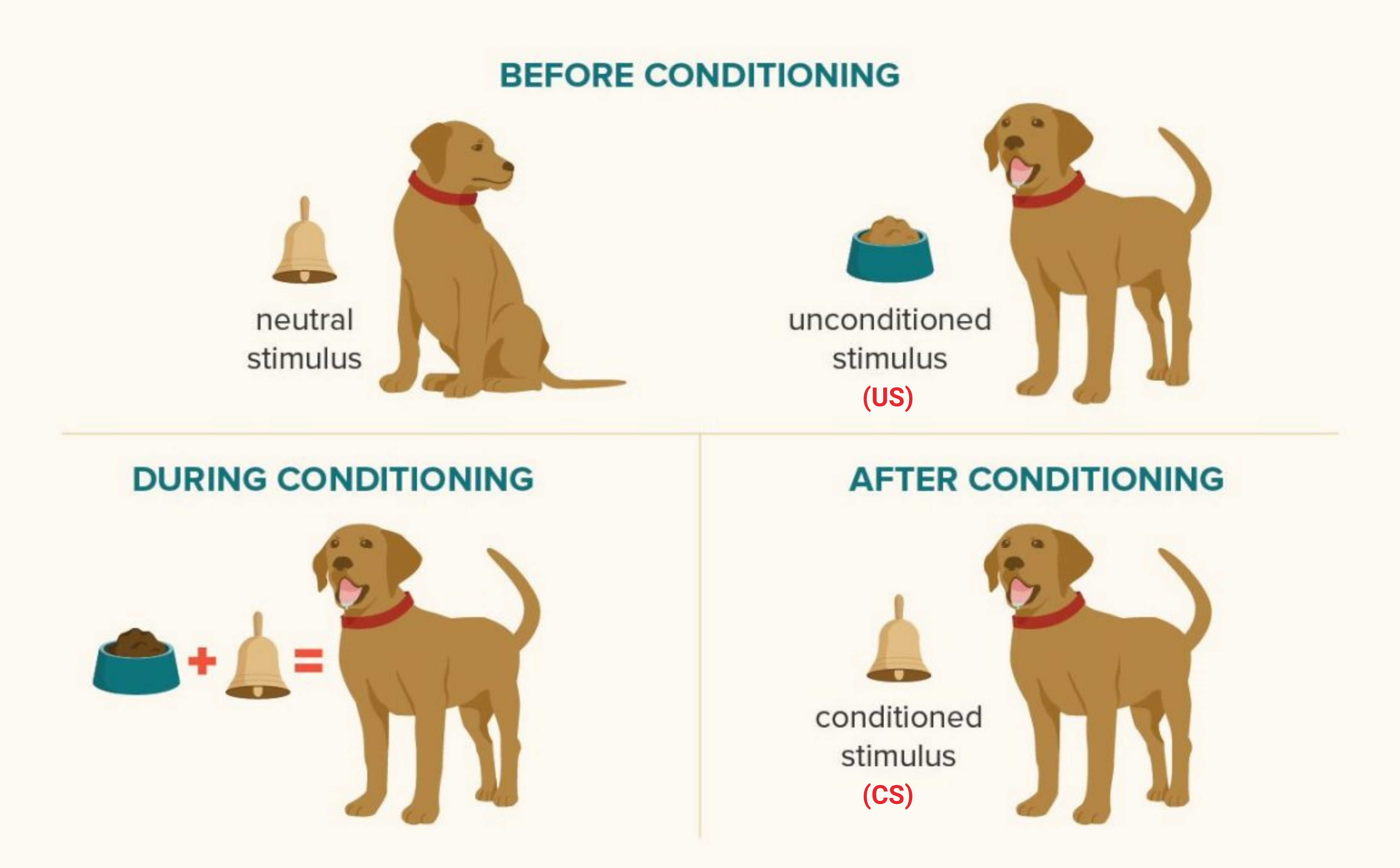
- Update predictions to match later, more accurate, predictions about the future before the final outcome is known.

$$P_{new} = (1-\alpha)P_{old} + \alpha(P_{correct})$$

$$\begin{aligned} P_{new} &= (1-\alpha)P_{old} + \alpha(P_{better}) \\ &= P_{old} + \alpha(P_{better} - P_{old}) \end{aligned}$$

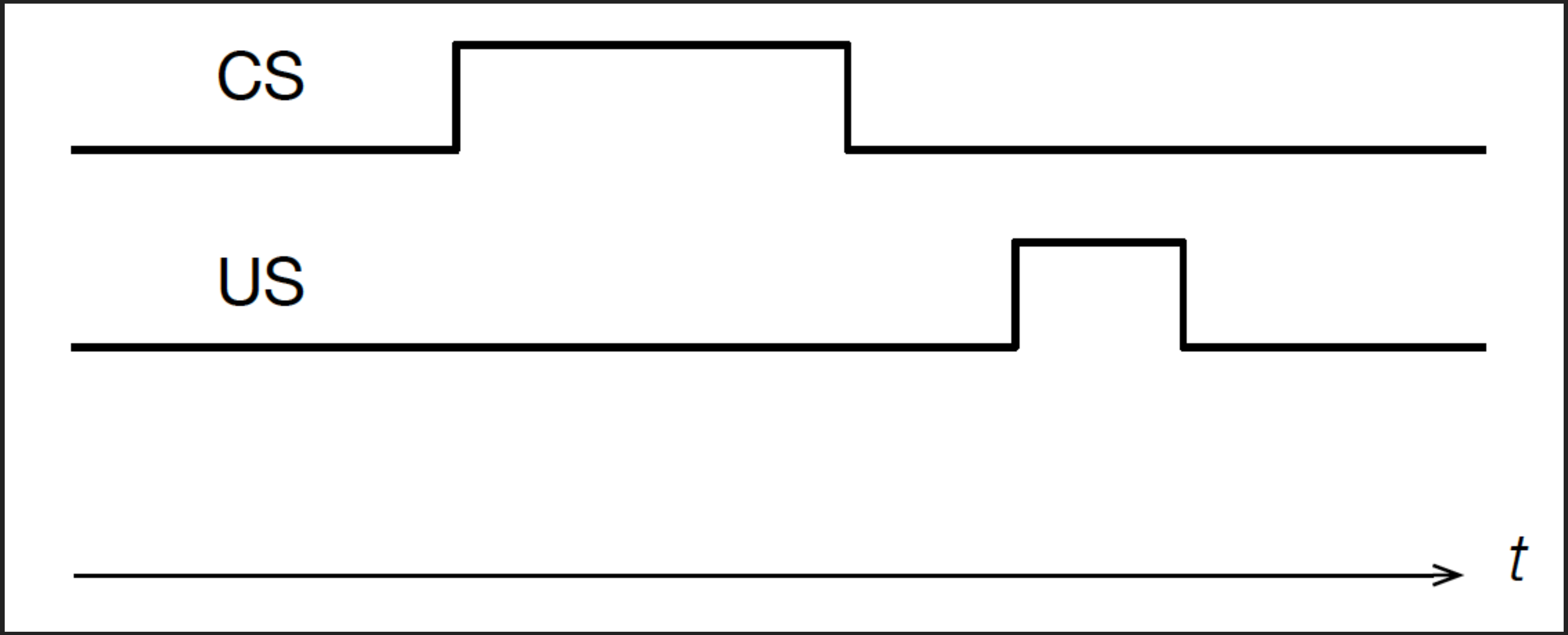
$$V_{new}(s) = V_{old}(s) + \alpha(\underbrace{R + V_{old}(s') - V_{old}(s)}_{\delta})$$

CLASSICAL CONDITIONING

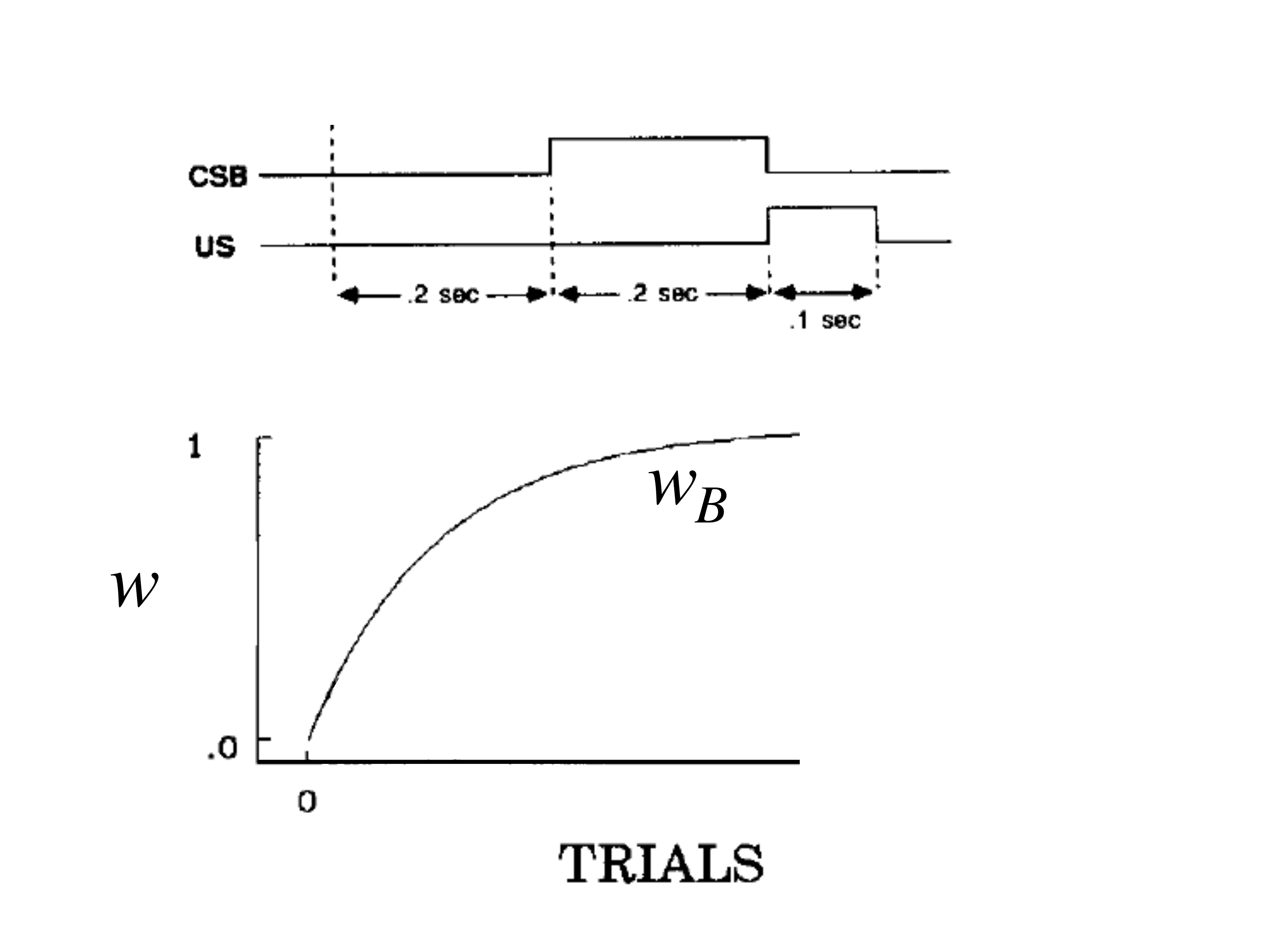


Credits: [Healthline]

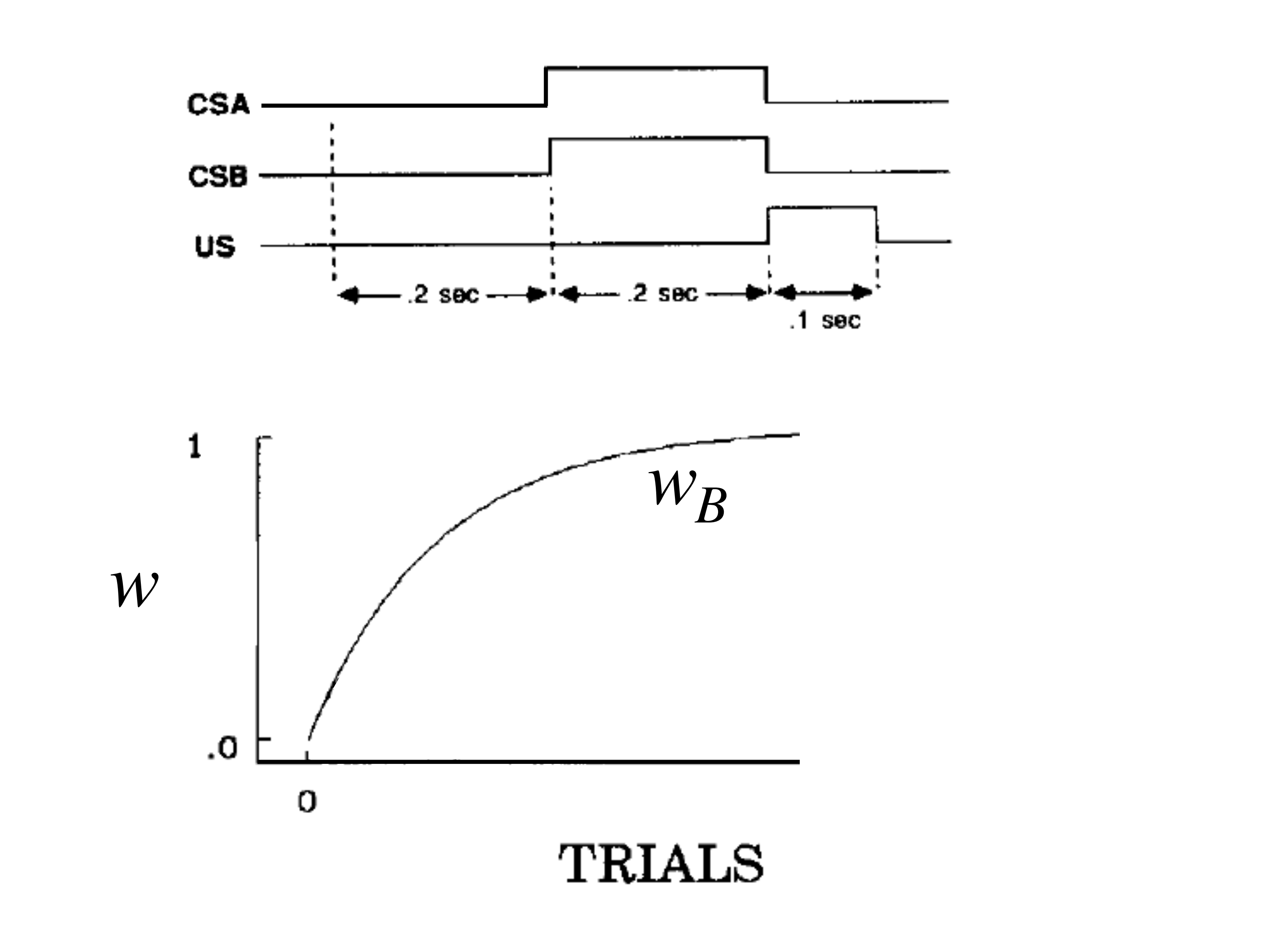
CLASSICAL CONDITIONING



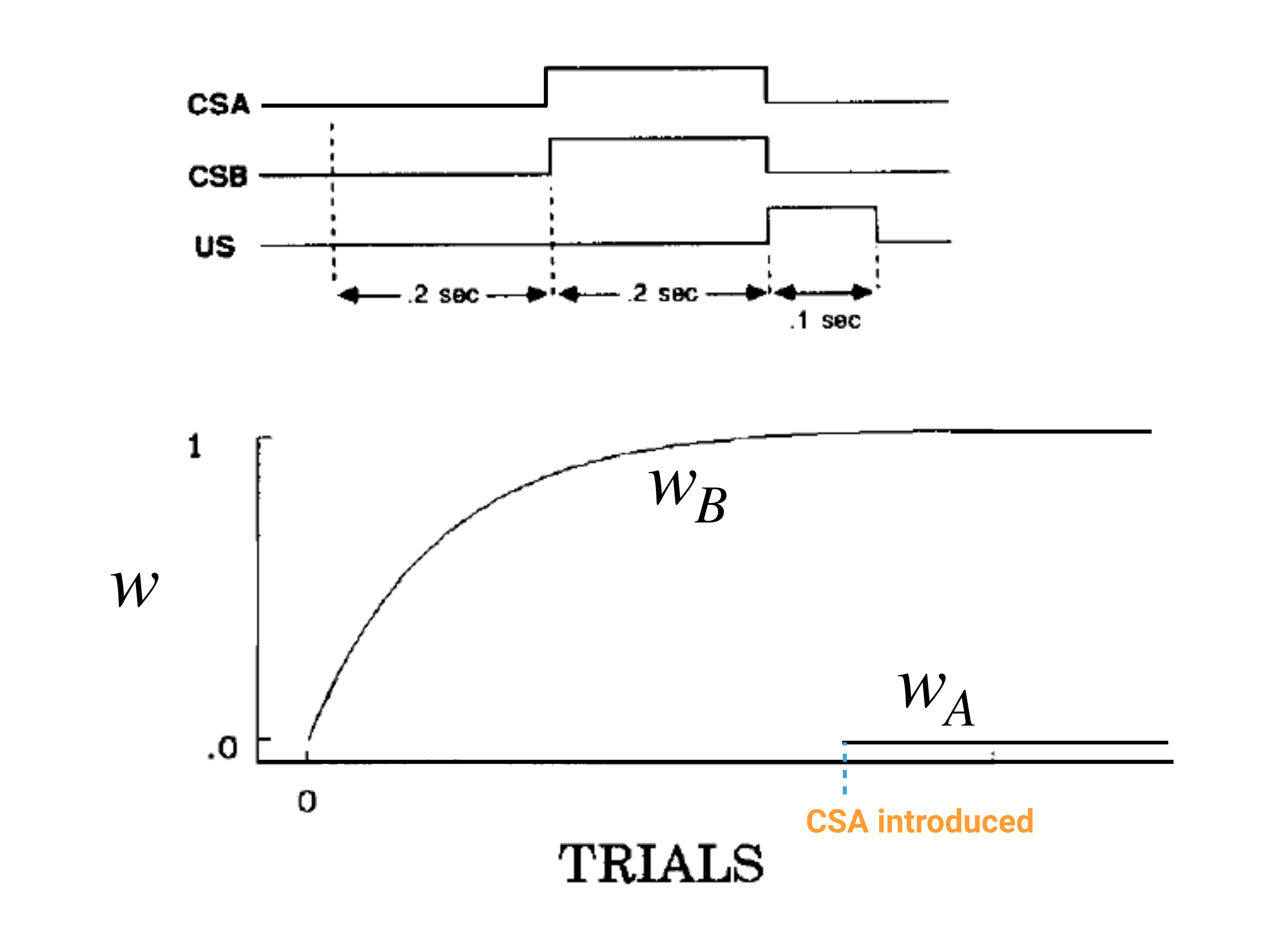
BLOCKING



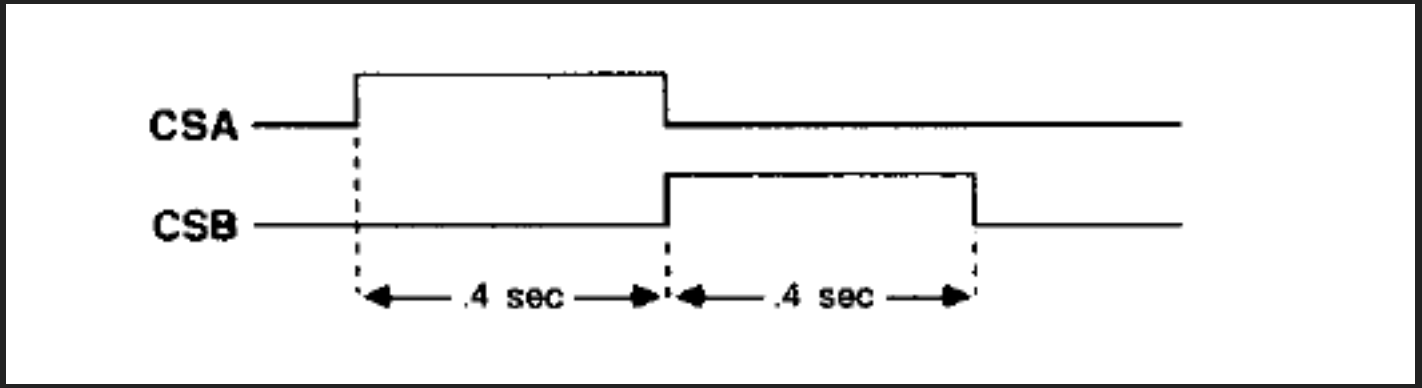
BLOCKING



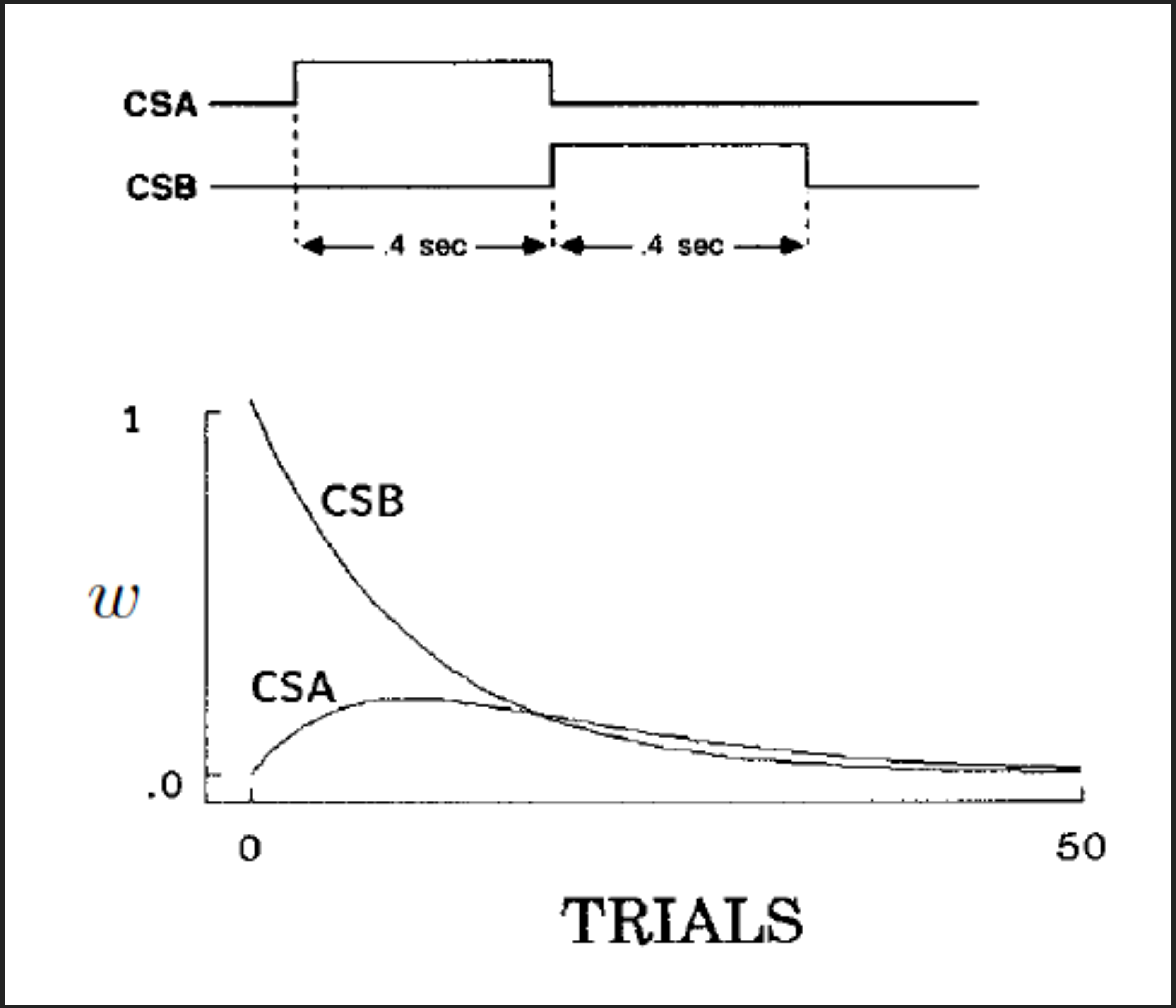
BLOCKING



HIGHER-ORDER CONDITIONING



HIGHER-ORDER CONDITIONING

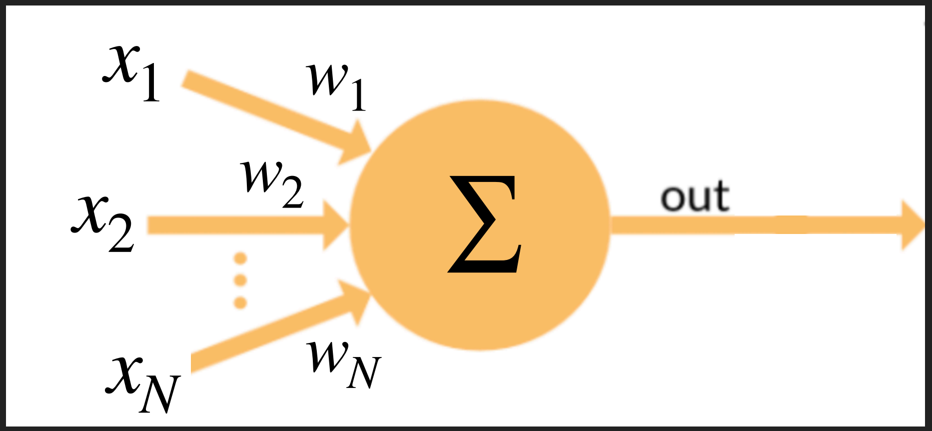


THE TD MODEL

$$v_w(s) = \sum_i w_i x_i(s) = w^T x(s)$$

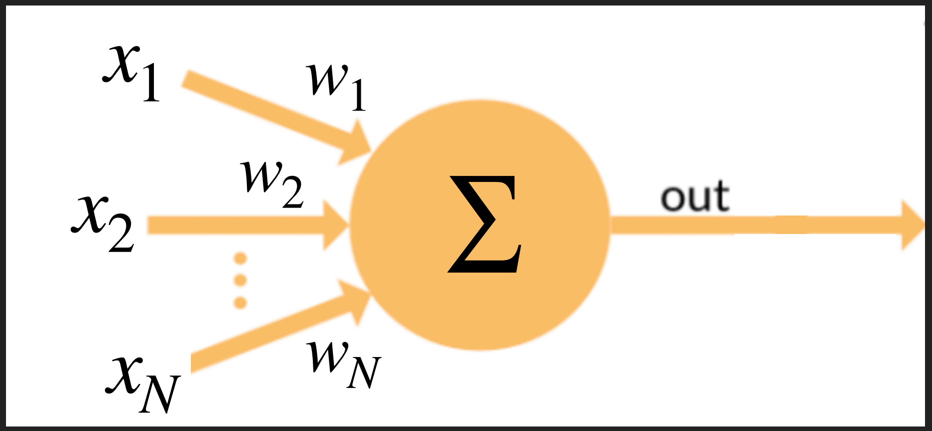
THE TD MODEL

$$v_w(s) = \sum_i w_i x_i(s) = w^T x(s)$$



THE TD MODEL

$$v_w(s) = \sum_i w_i x_i(s) = w^T x(s)$$

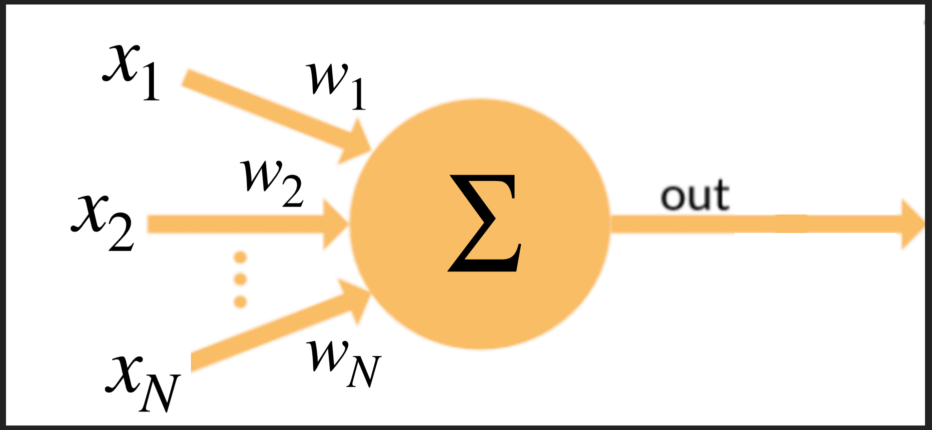


$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

$$w_{t+1} = w_t + \alpha \delta_t x_t$$

THE TD MODEL

$$v_w(s) = \sum_i w_i x_i(s) = w^T x(s)$$



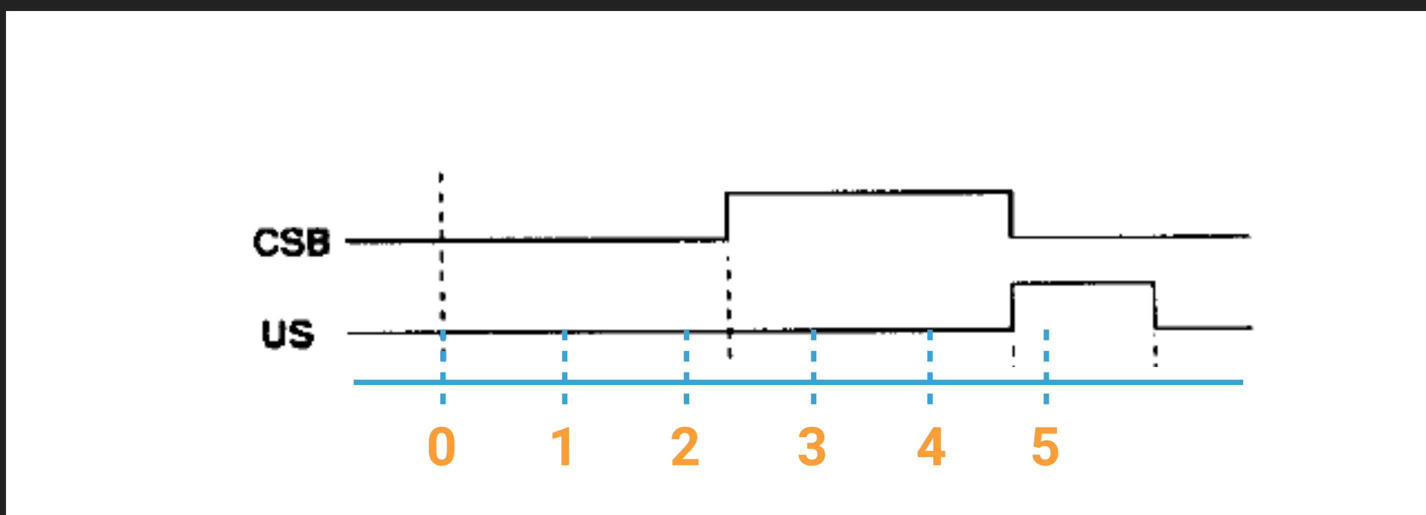
$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

$$w_{t+1} = w_t + \alpha \delta_t x_t$$

- The TD model explains blocking and higher-order conditioning

THE TD MODEL EXPLAINS BLOCKING

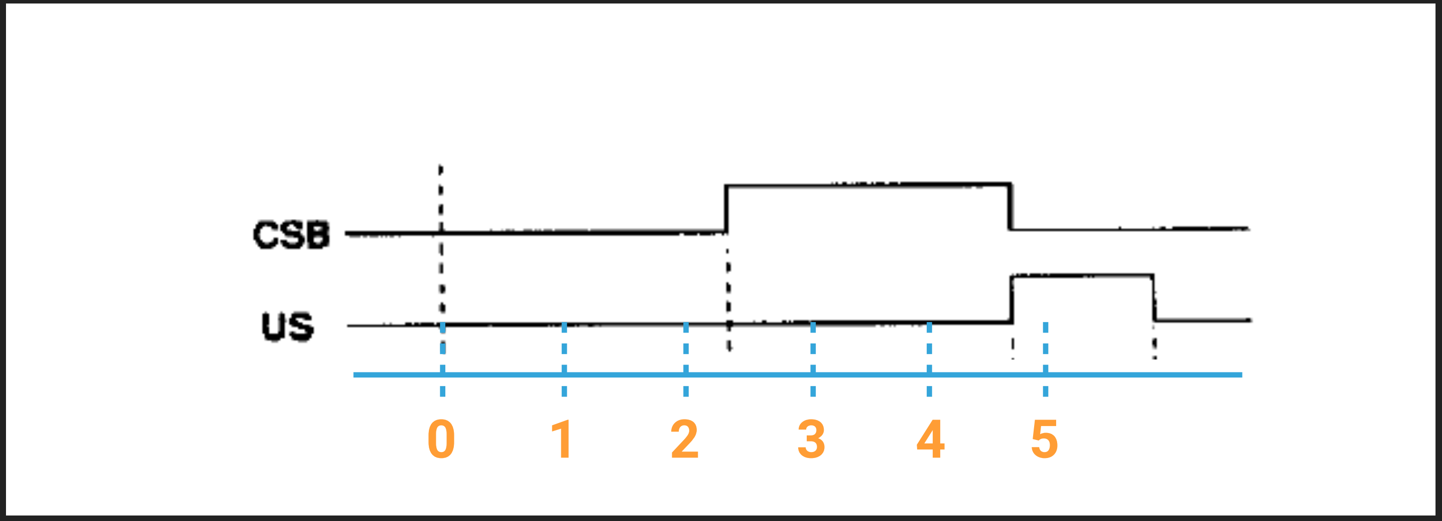
$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



When only CSB is trained:

THE TD MODEL EXPLAINS BLOCKING

$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



When only CSB is trained:

$$\delta_4 = R_5 + v_w(S_5) - v_w(S_4)$$

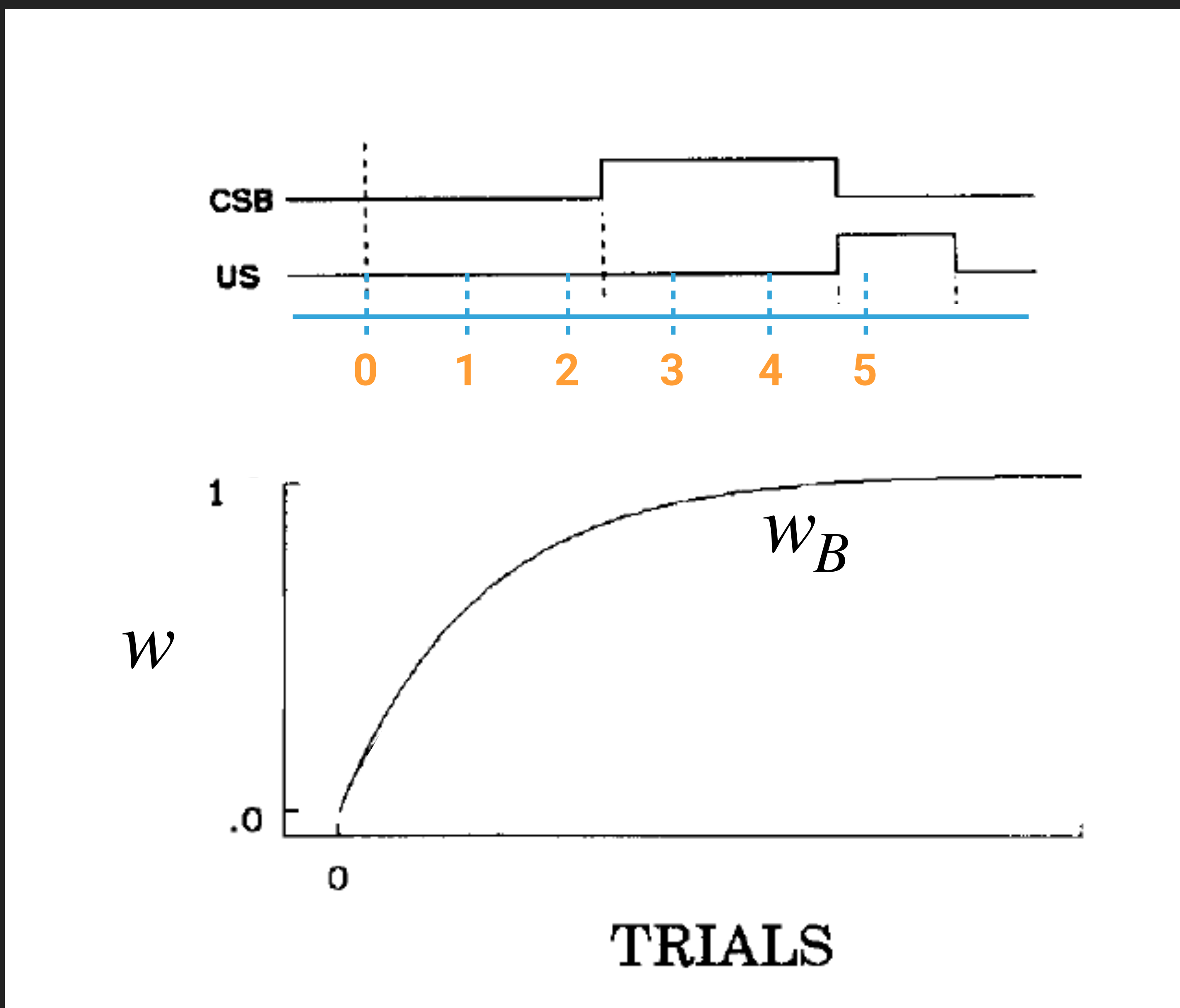
$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_B \cdot 1)$$

$$\implies w_B = 1$$

THE TD MODEL EXPLAINS BLOCKING

$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



When only CSB is trained:

$$\delta_4 = R_5 + v_w(S_5) - v_w(S_4)$$

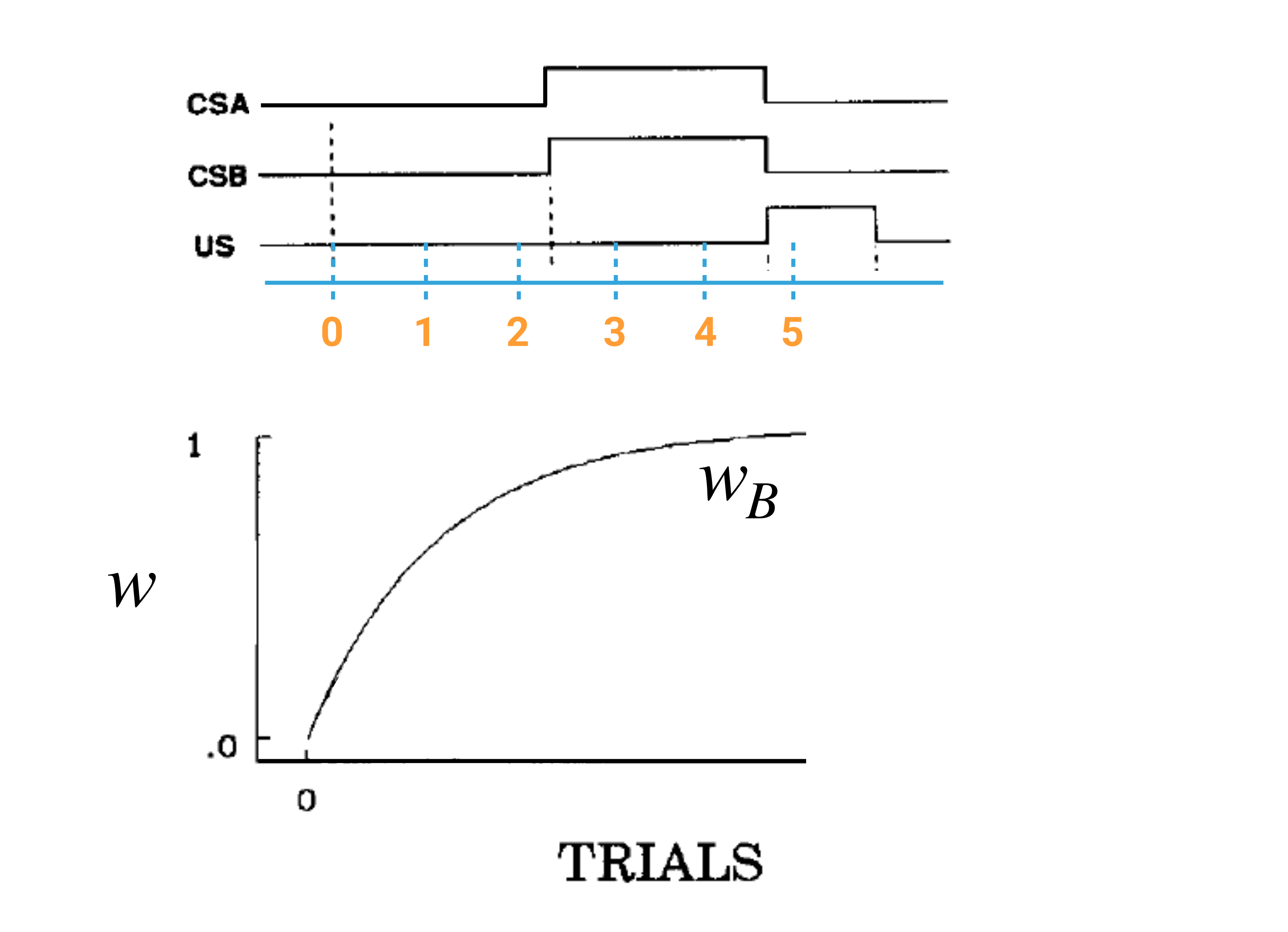
$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_B \cdot 1)$$

$$\implies w_B = 1$$

THE TD MODEL EXPLAINS BLOCKING

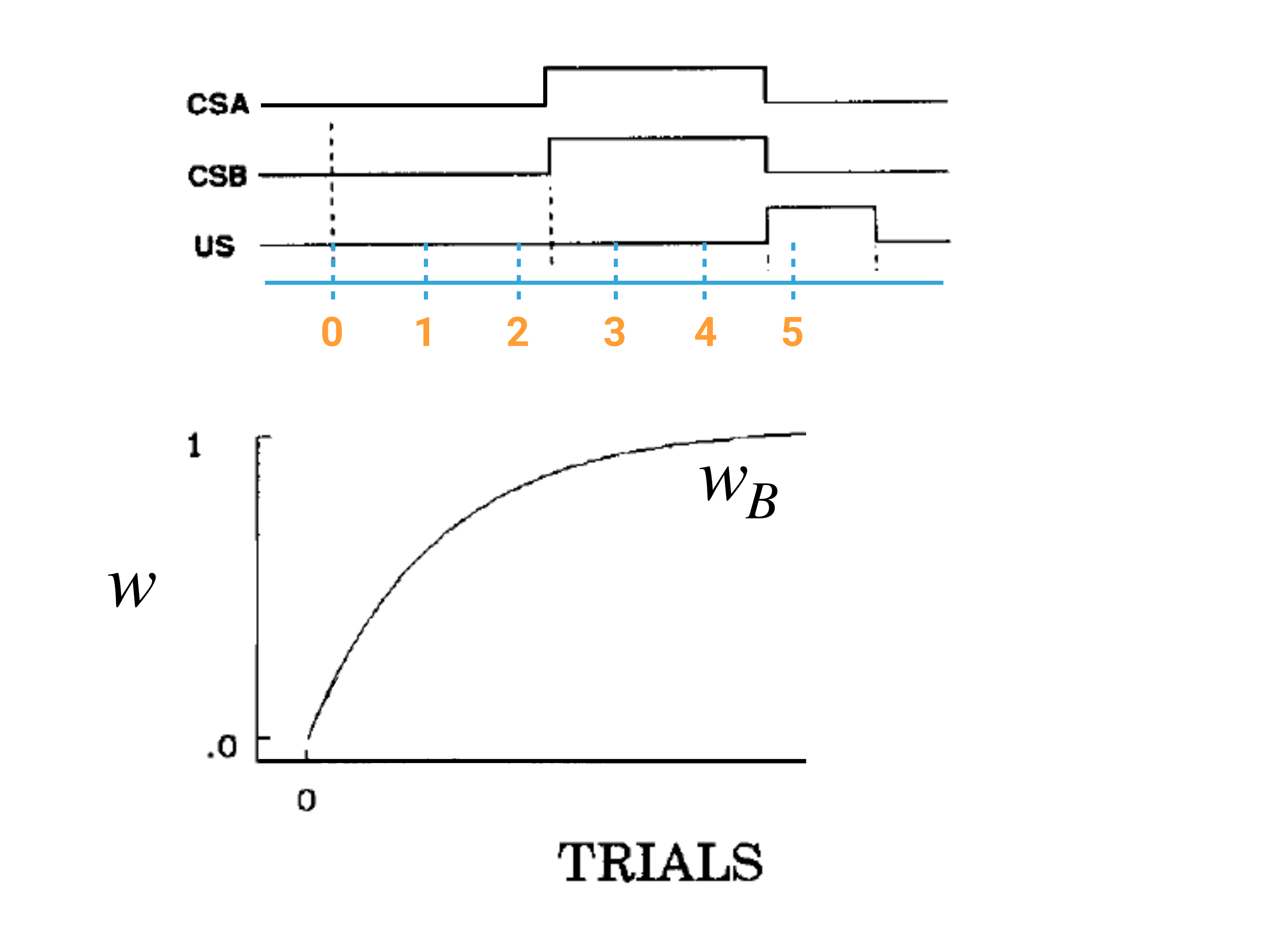
$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



After CSA is added:

THE TD MODEL EXPLAINS BLOCKING

$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



After CSA is added:

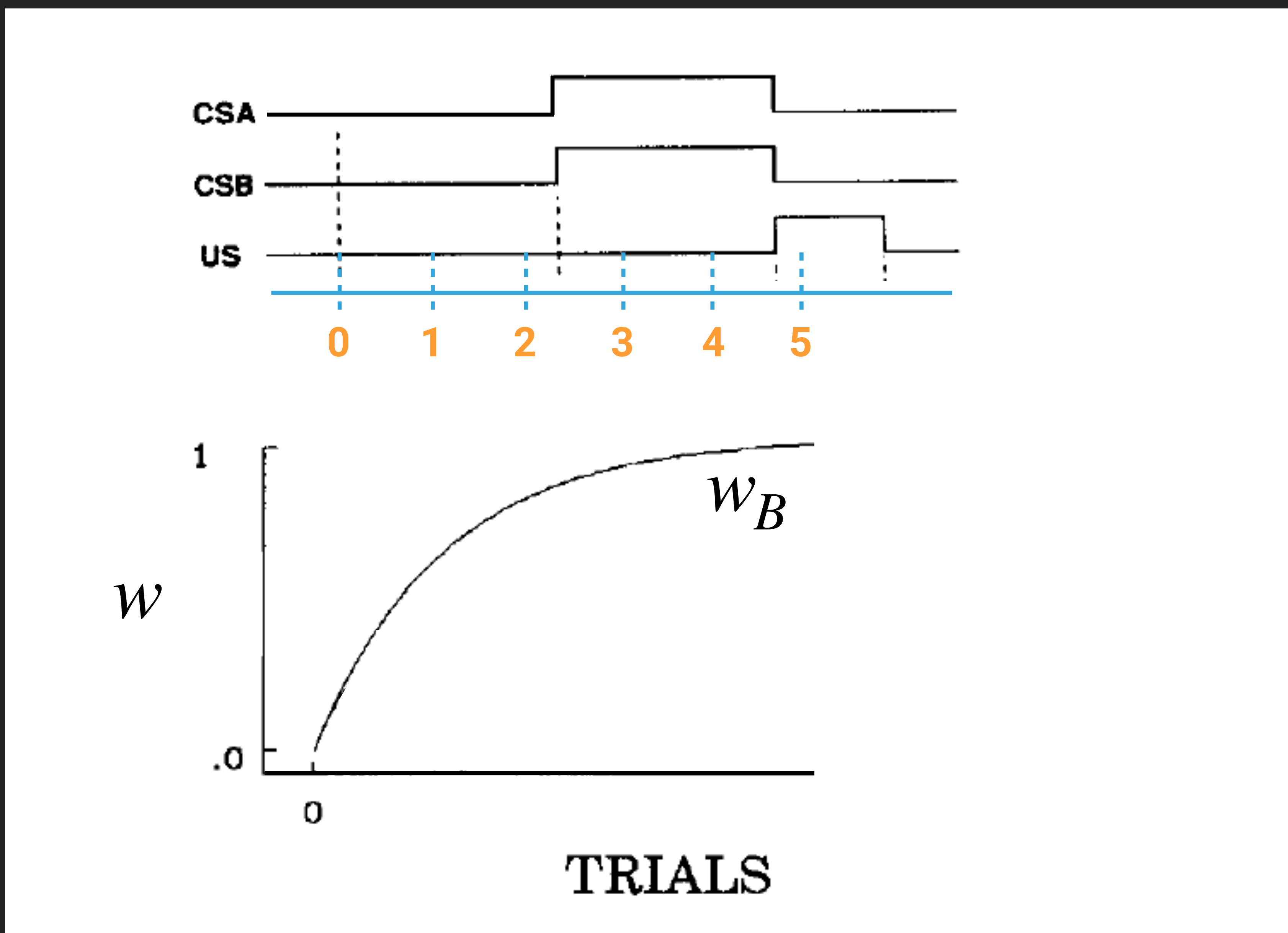
$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

$$\Rightarrow \; w_A + w_B = 1$$

THE TD MODEL EXPLAINS BLOCKING

$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



After CSA is added:

$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

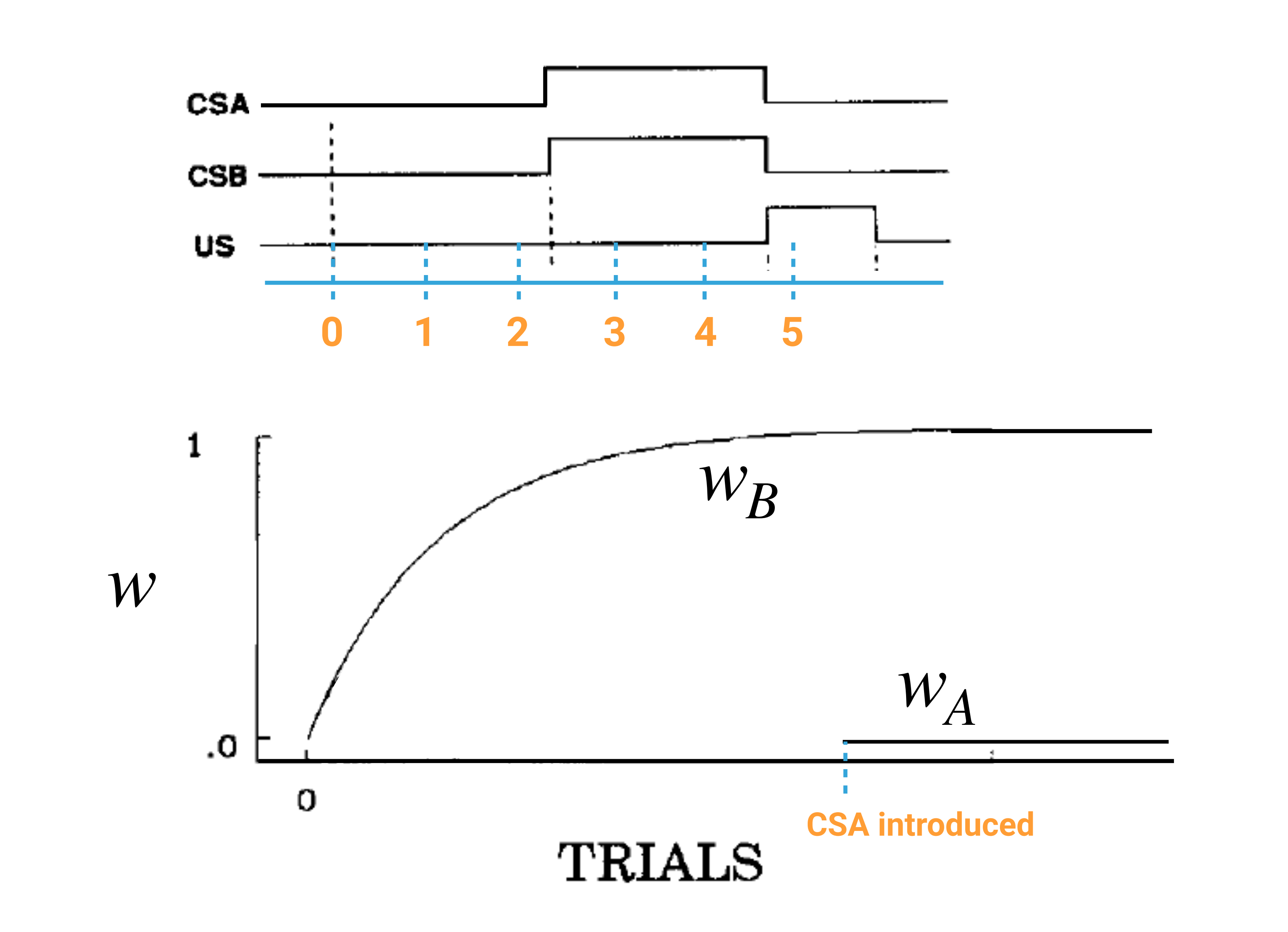
$$\implies w_A + w_B = 1$$

But $w_B = 1$

$$\implies w_A = 0$$

THE TD MODEL EXPLAINS BLOCKING

$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$



After CSA is added:

$$\delta_4 = 1 + 0 - w^T x_4$$

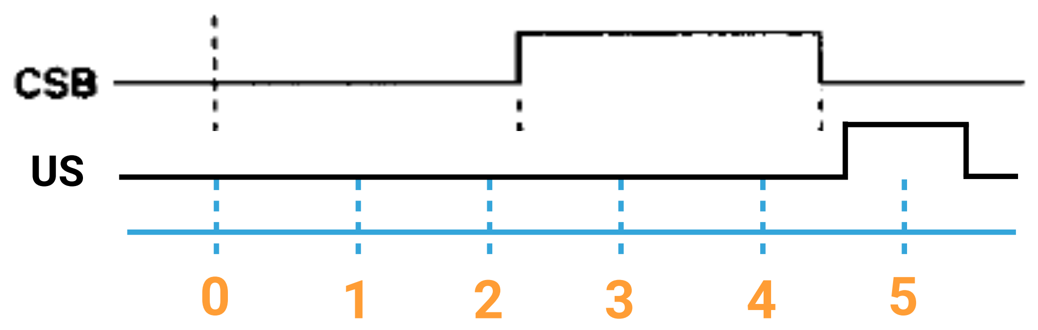
$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

$$\Rightarrow \; w_A + w_B = 1$$

But $w_B = 1$

$$\Longrightarrow \; w_A = 0$$

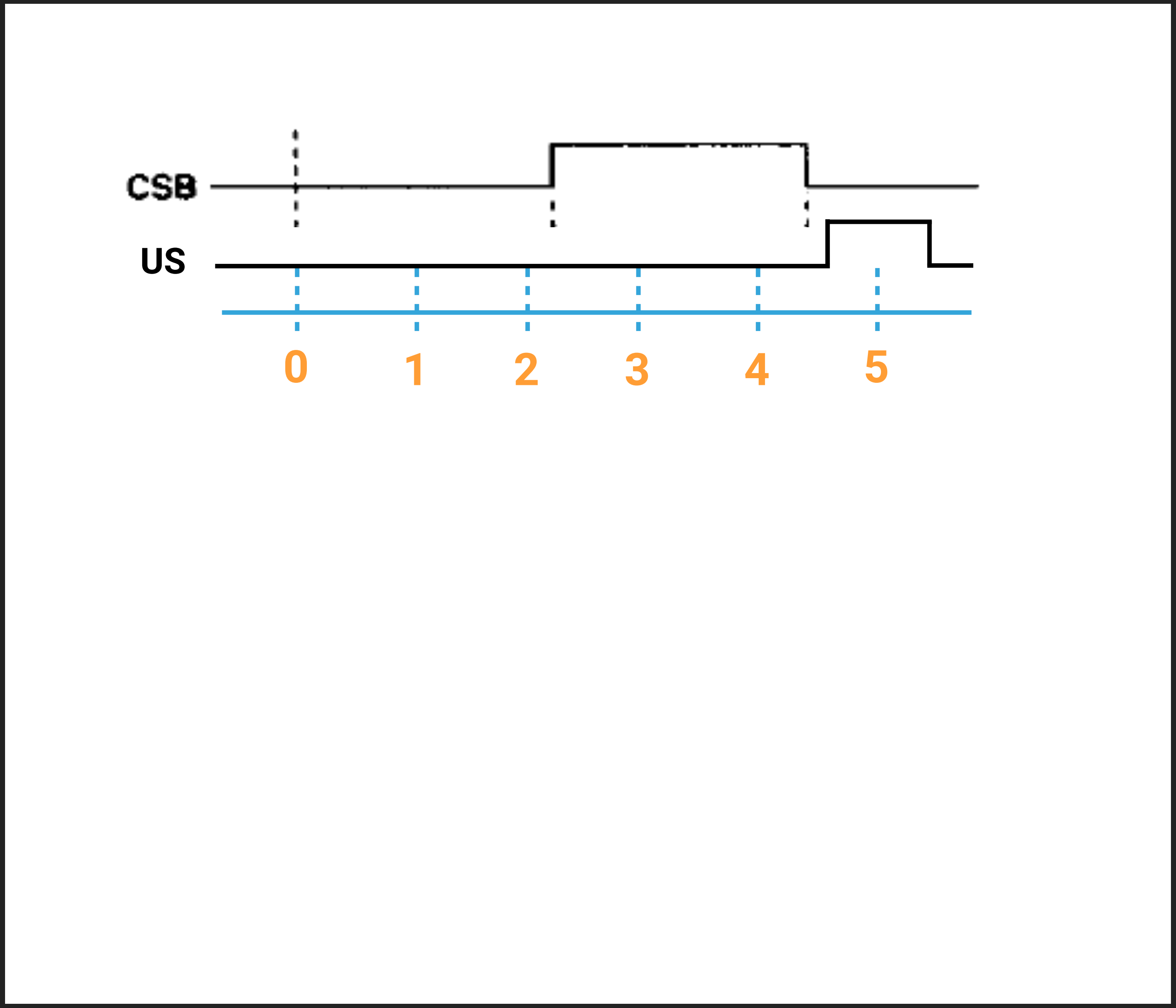
THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

When only CSB is trained:

THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING

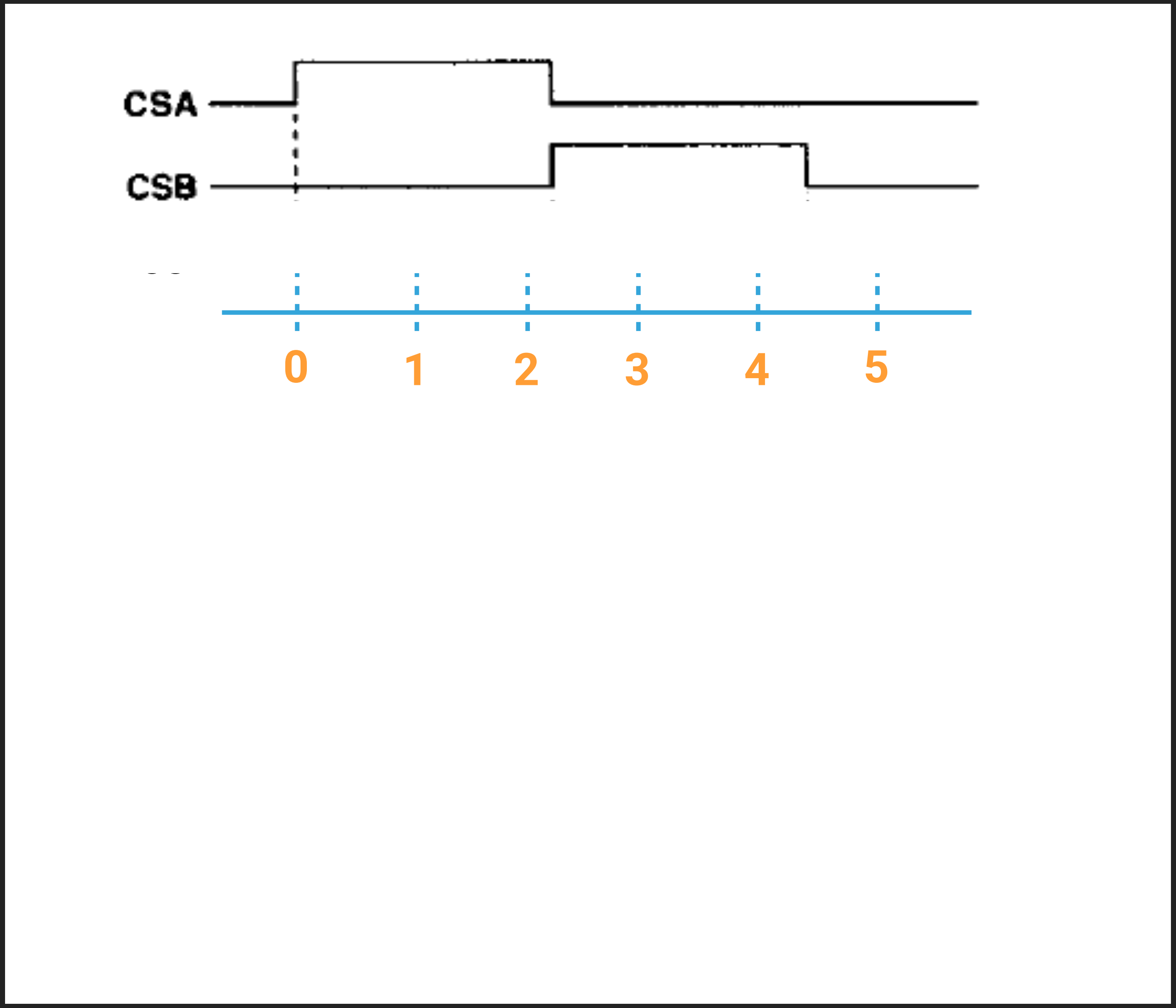


$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

When only CSB is trained:

$$\delta_4 = 1 + 0 - w^T x_4$$
$$0 = 1 + 0 - (w_B \cdot 1)$$
$$\implies w_B = 1$$

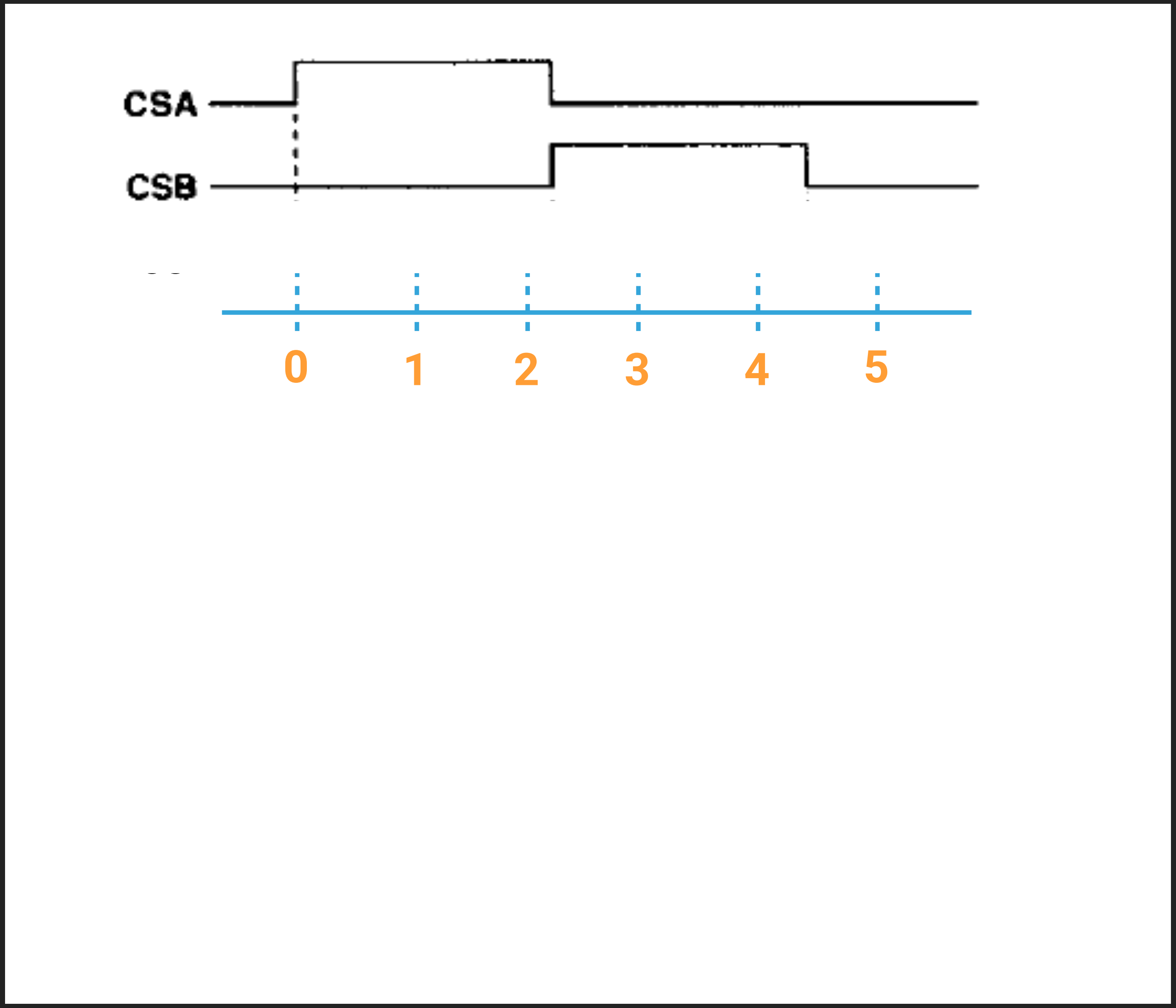
THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

After CSA introduced and US removed:

THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

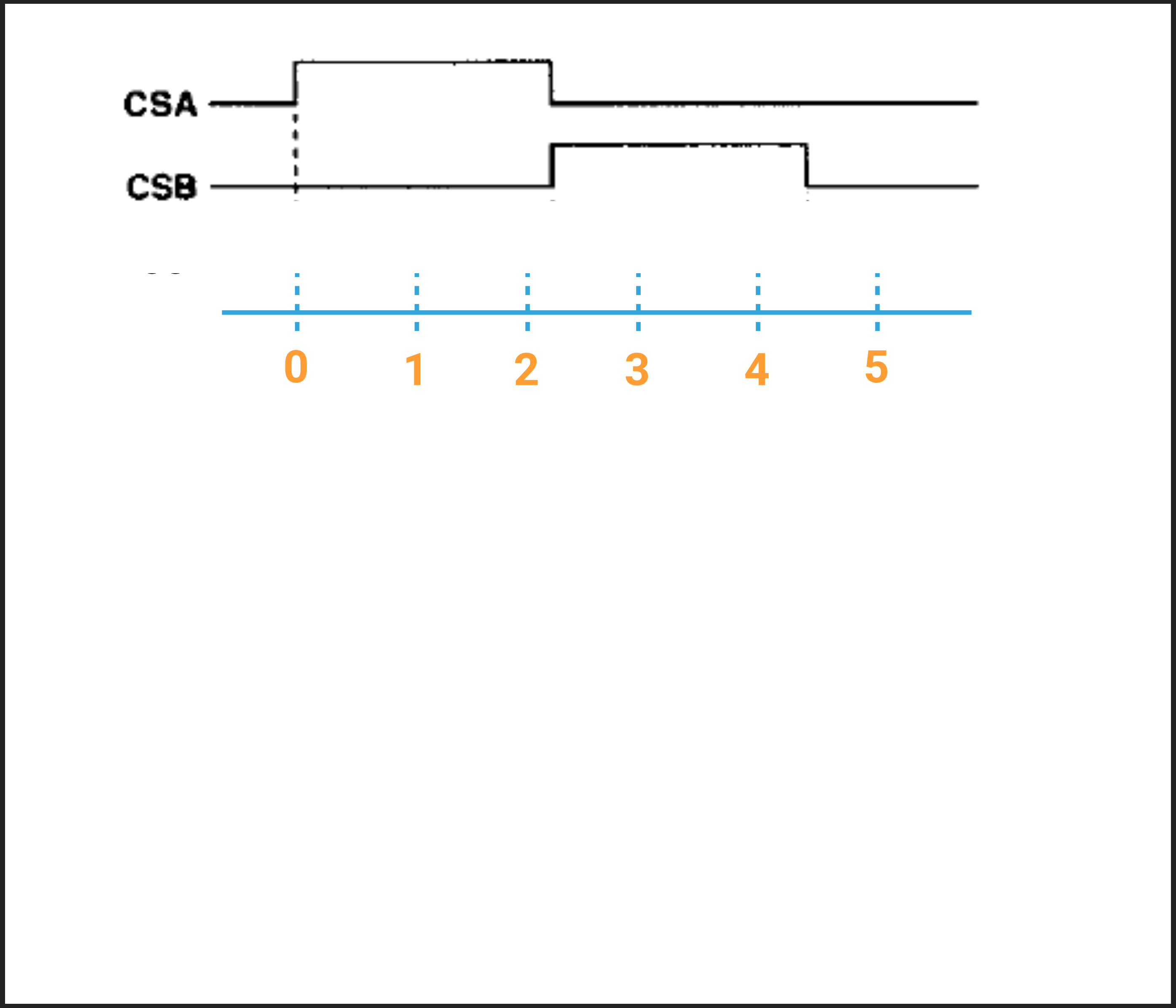
After CSA introduced and US removed:

$$\delta_2 = 0 + w^T x_3 - w^T x_2$$

$$0 = 0 + (w_B 1) - (w_A 1)$$

$$\implies w_A = w_B$$

THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING



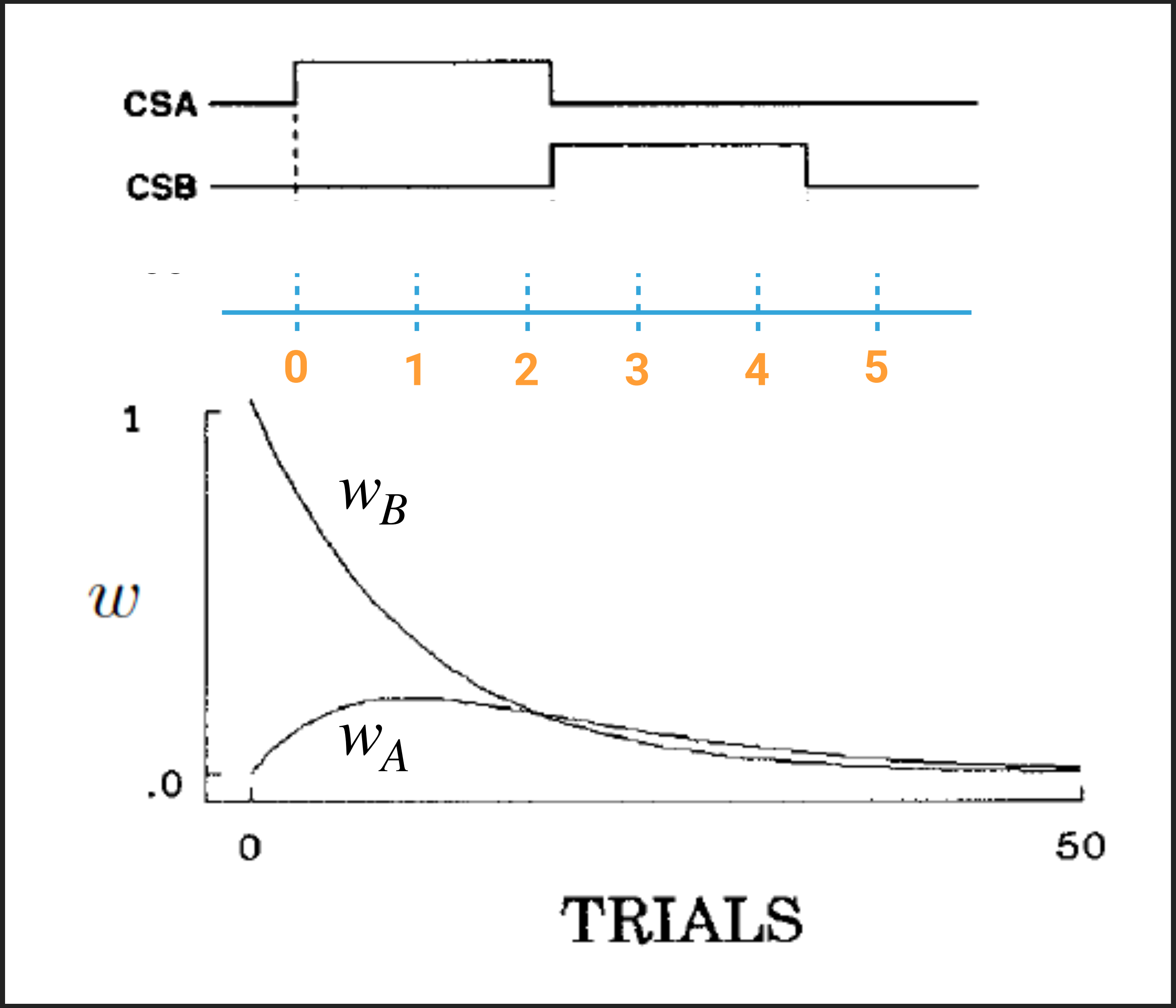
$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

After CSA introduced and US removed:

$$\delta_2 = 0 + w^T x_3 - w^T x_2$$
$$0 = 0 + (w_B 1) - (w_A 1)$$
$$\implies w_A = w_B$$

$$\delta_4 = 0 + 0 - w^T x_4$$
$$0 = 0 + 0 - (w_B 1)$$
$$\implies w_B = 0$$

THE TD MODEL EXPLAINS HIGHER-ORDER CONDITIONING



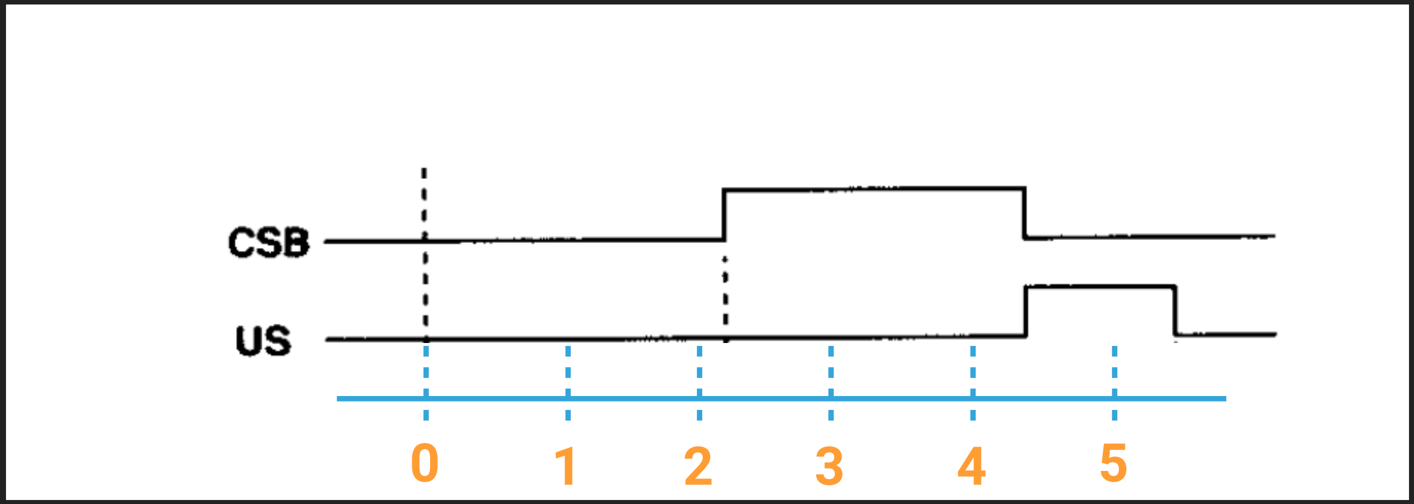
$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

After CSA introduced and US removed:

$$\delta_2 = 0 + w^T x_3 - w^T x_2$$
$$0 = 0 + (w_B 1) - (w_A 1)$$
$$\implies w_A = w_B$$

$$\delta_4 = 0 + 0 - w^T x_4$$
$$0 = 0 + 0 - (w_B 1)$$
$$\implies w_B = 0$$

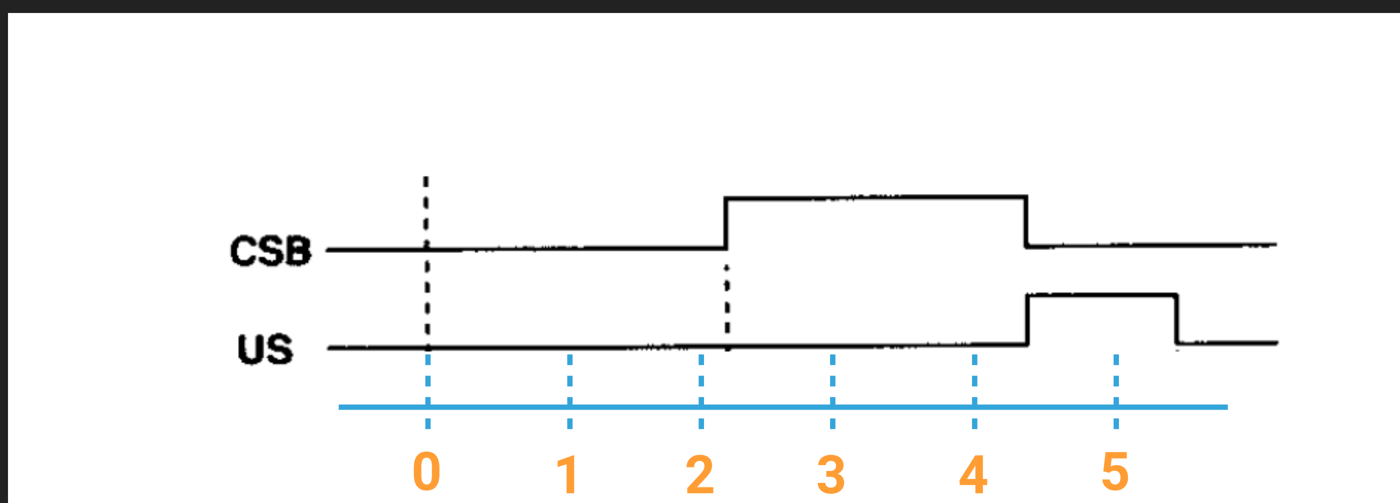
THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Only CSB is trained:

THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING

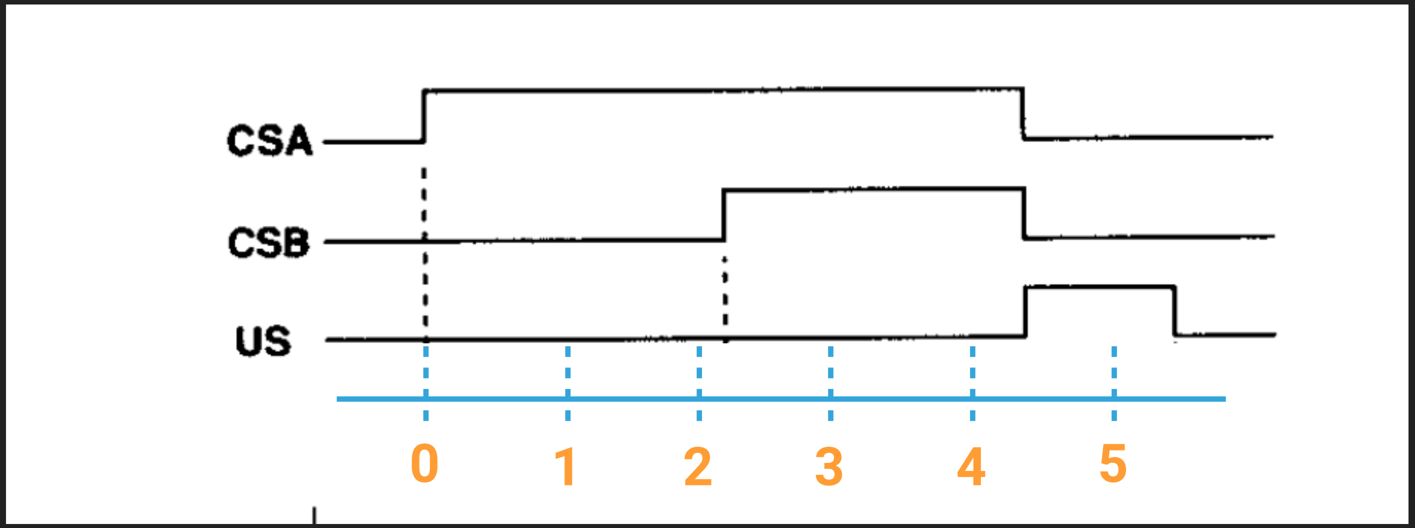


$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Only CSB is trained:

$$\delta_4 = 1 + 0 - w^T x_4$$
$$0 = 1 + 0 - (w_B \cdot 1)$$
$$\implies w_B = 1$$

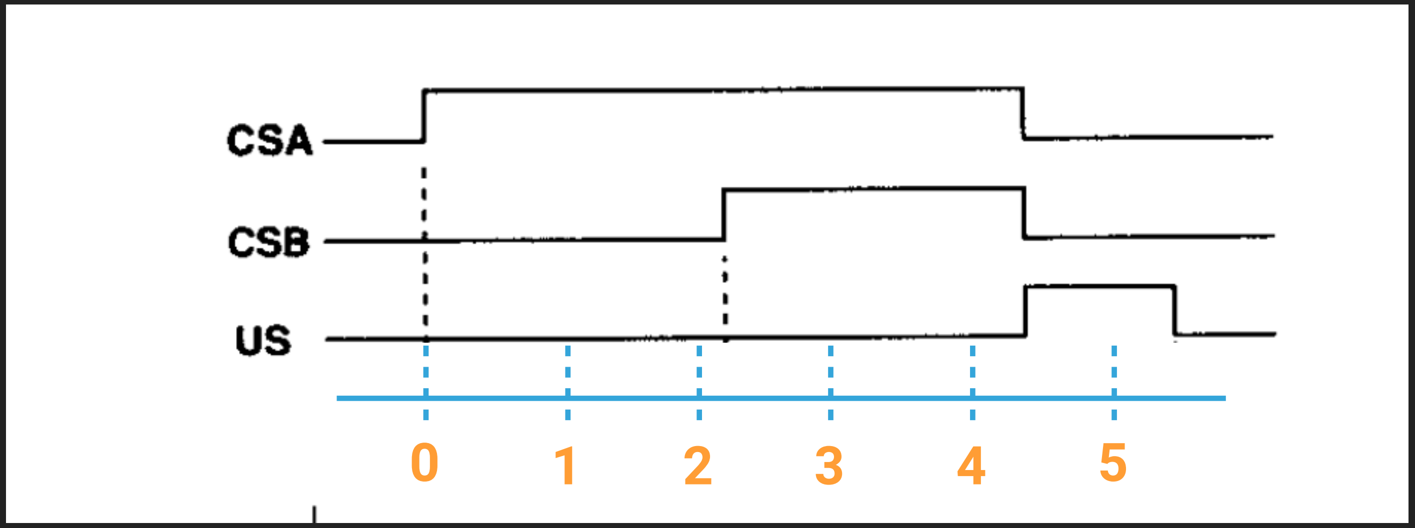
THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Now CSA is introduced before the onset of CSB:

THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

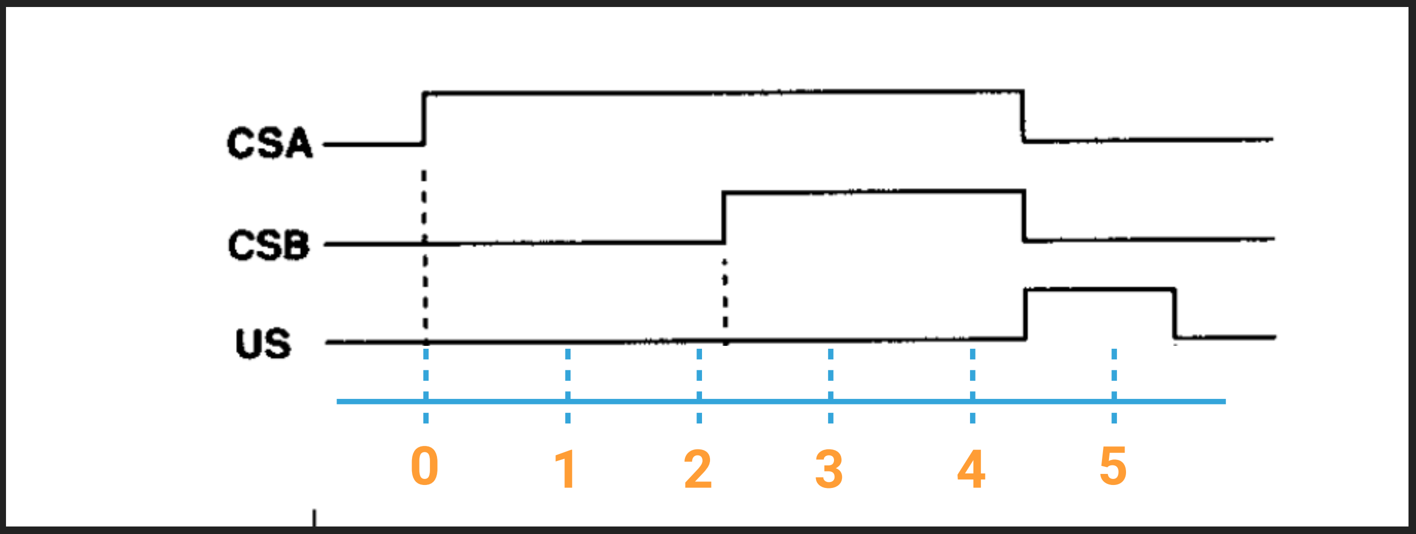
Now CSA is introduced before the onset of CSB:

$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

$$\Rightarrow \quad w_A + w_B = 1$$

THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Now CSA is introduced before the onset of CSB:

$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

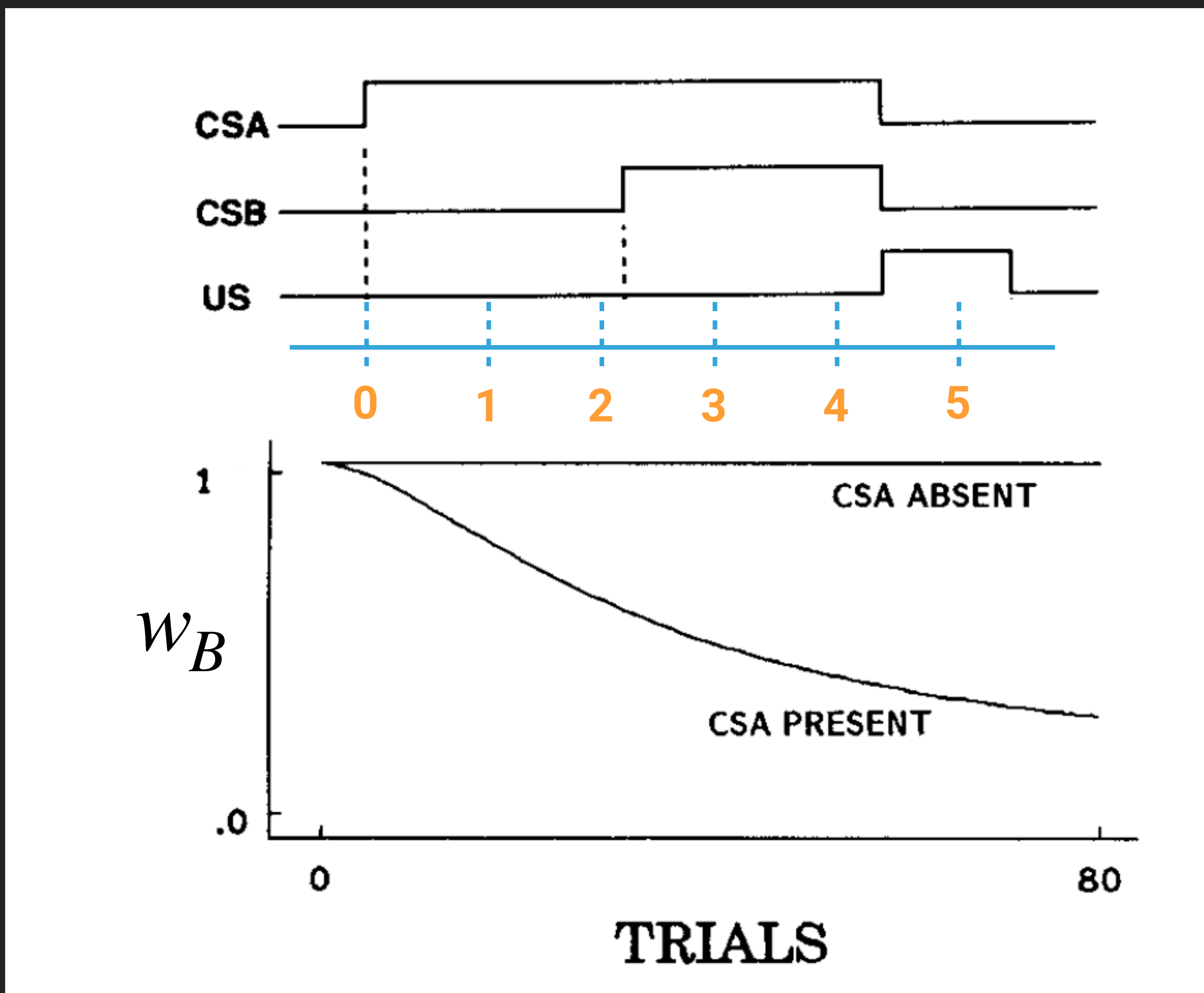
$$\Rightarrow \quad w_A + w_B = 1$$

$$\delta_2 = 0 + w^T x_3 - w^T x_2$$

$$0 = 0 + (w_B 1 + w_A 1) - (w_A 1)$$

$$\Rightarrow \quad w_B = 0$$

THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Now CSA is introduced before the onset of CSB:

$$\delta_4 = 1 + 0 - w^T x_4$$

$$0 = 1 + 0 - (w_A 1 + w_B 1)$$

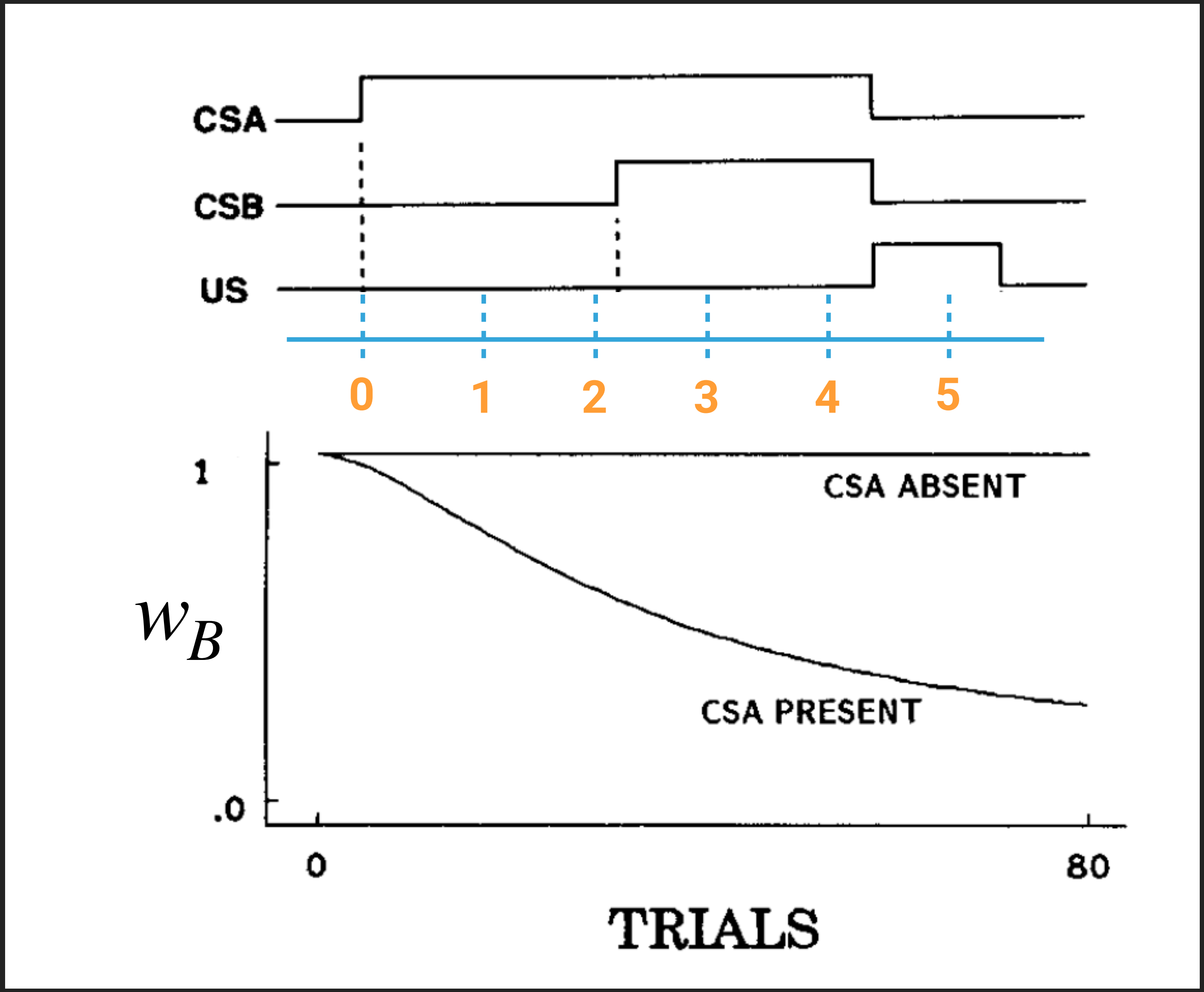
$$\implies w_A + w_B = 1$$

$$\delta_2 = 0 + w^T x_3 - w^T x_2$$

$$0 = 0 + (w_B 1 + w_A 1) - (w_A 1)$$

$$\implies w_B = 0$$

THE TD MODEL PREDICTS THE REVERSAL OF BLOCKING



$$\delta_t = R_{t+1} + v_w(S_{t+1}) - v_w(S_t)$$

Now CSA is introduced before the onset of CSB:

$$\delta_4 = 1 + 0 - w^T x_4$$
$$0 = 1 + 0 - (w_A 1 + w_B 1)$$
$$\implies w_A + w_B = 1$$

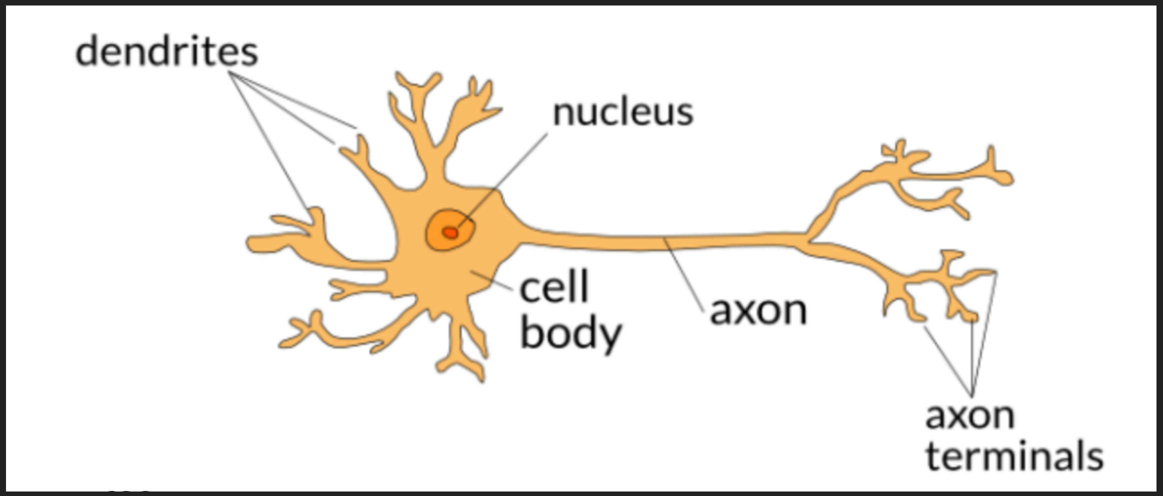
$$\delta_2 = 0 + w^T x_3 - w^T x_2$$
$$0 = 0 + (w_B 1 + w_A 1) - (w_A 1)$$
$$\implies w_B = 0$$

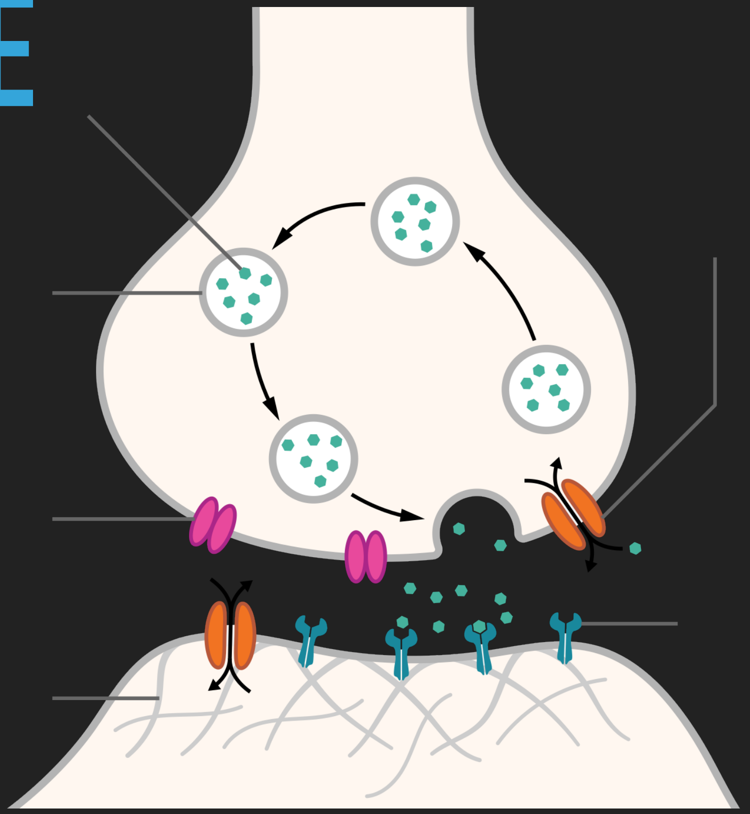
- This predicted phenomenon was later confirmed by Kehoe, Schreurs, and Graham (1987).

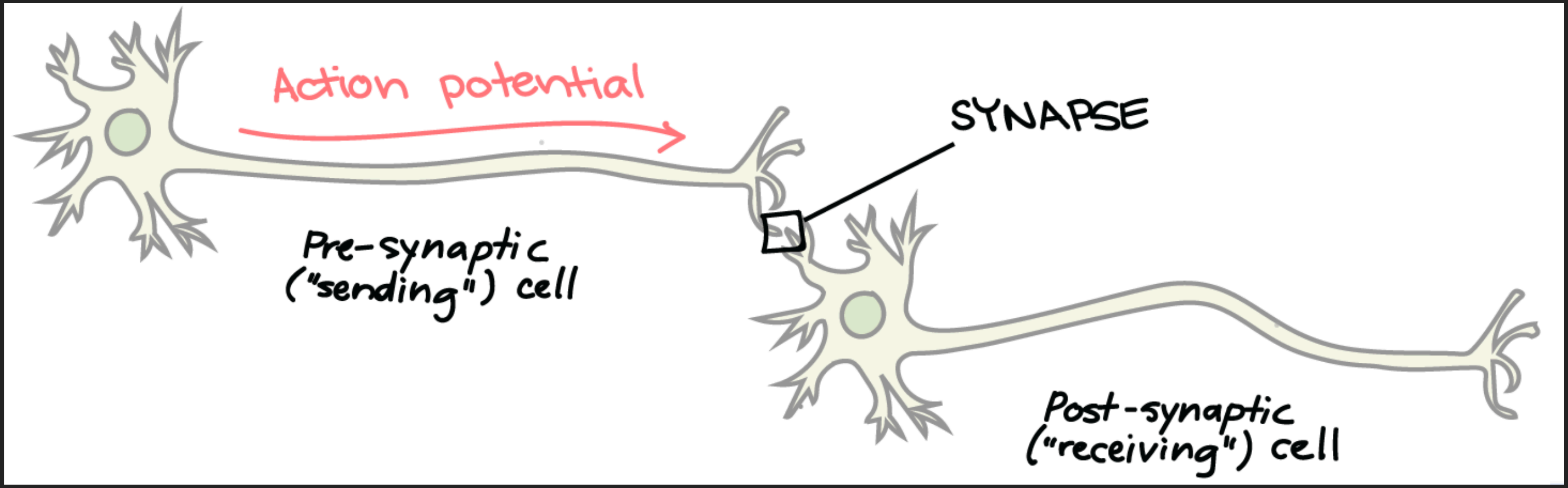
TAKEAWAY

The TD model not only explains a variety of psychological phenomena, but also predicted some new phenomena.

NEURONS, SYNAPSES, DOPAMINE







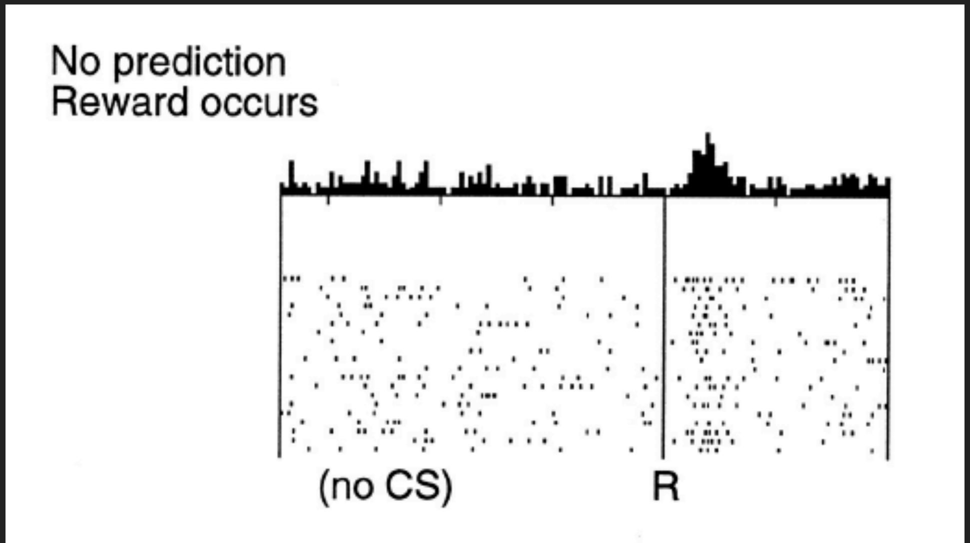
Credits: [Khan Academy](), [Wikipedia]()

THE REWARD PREDICTION ERROR HYPOTHESIS

One of the functions of the phasic activity of dopamine-producing neurons in mammals is to deliver an error between an old and a new estimate of expected future reward to target areas throughout the brain.

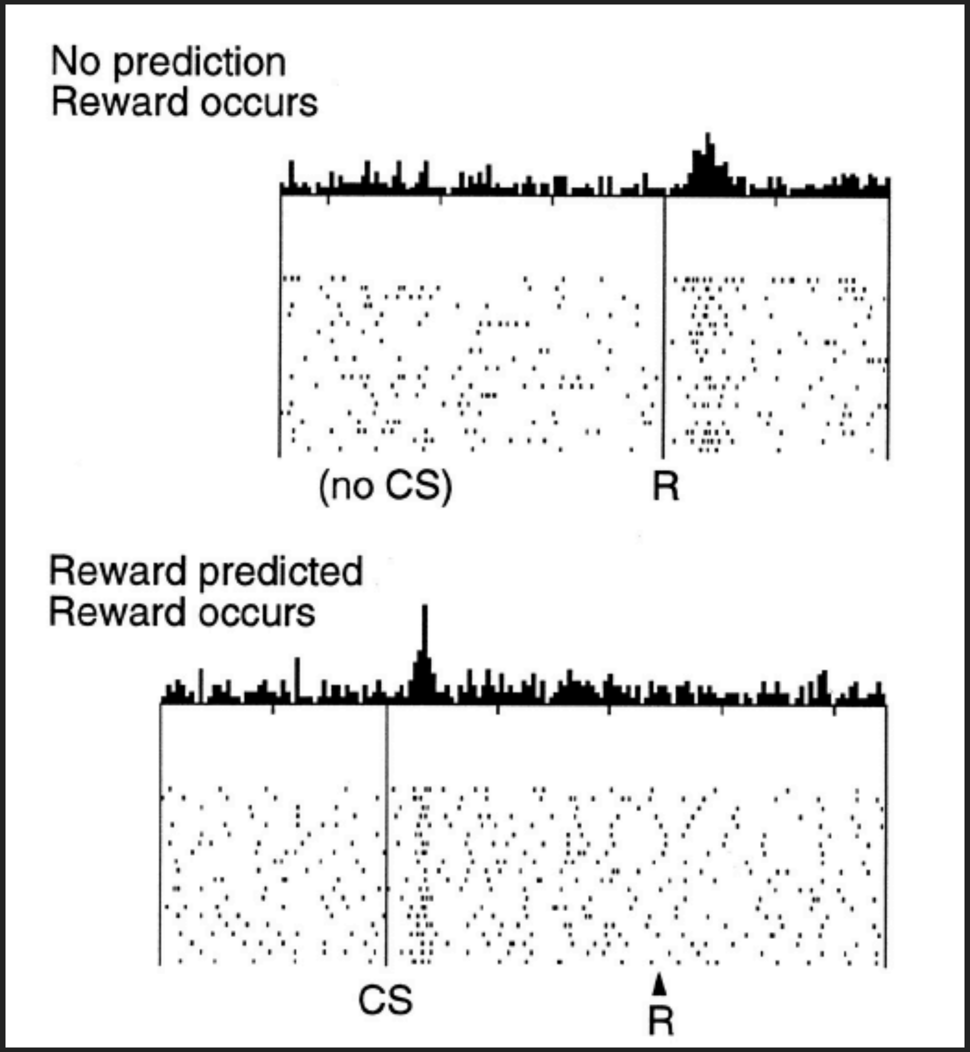
THE REWARD PREDICTION ERROR HYPOTHESIS

One of the functions of the phasic activity of dopamine-producing neurons in mammals is to deliver an error between an old and a new estimate of expected future reward to target areas throughout the brain.



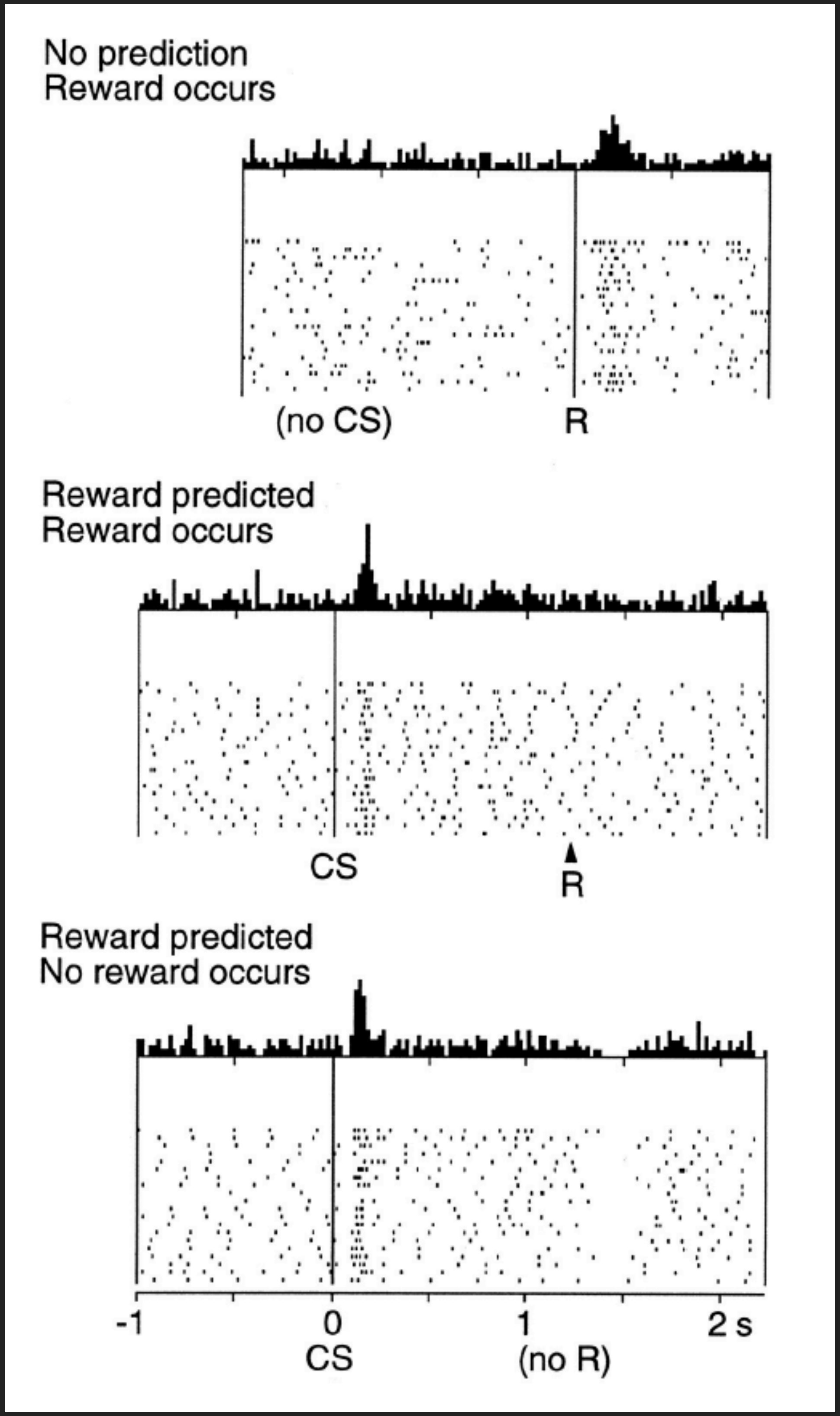
THE REWARD PREDICTION ERROR HYPOTHESIS

One of the functions of the phasic activity of dopamine-producing neurons in mammals is to deliver an error between an old and a new estimate of expected future reward to target areas throughout the brain.



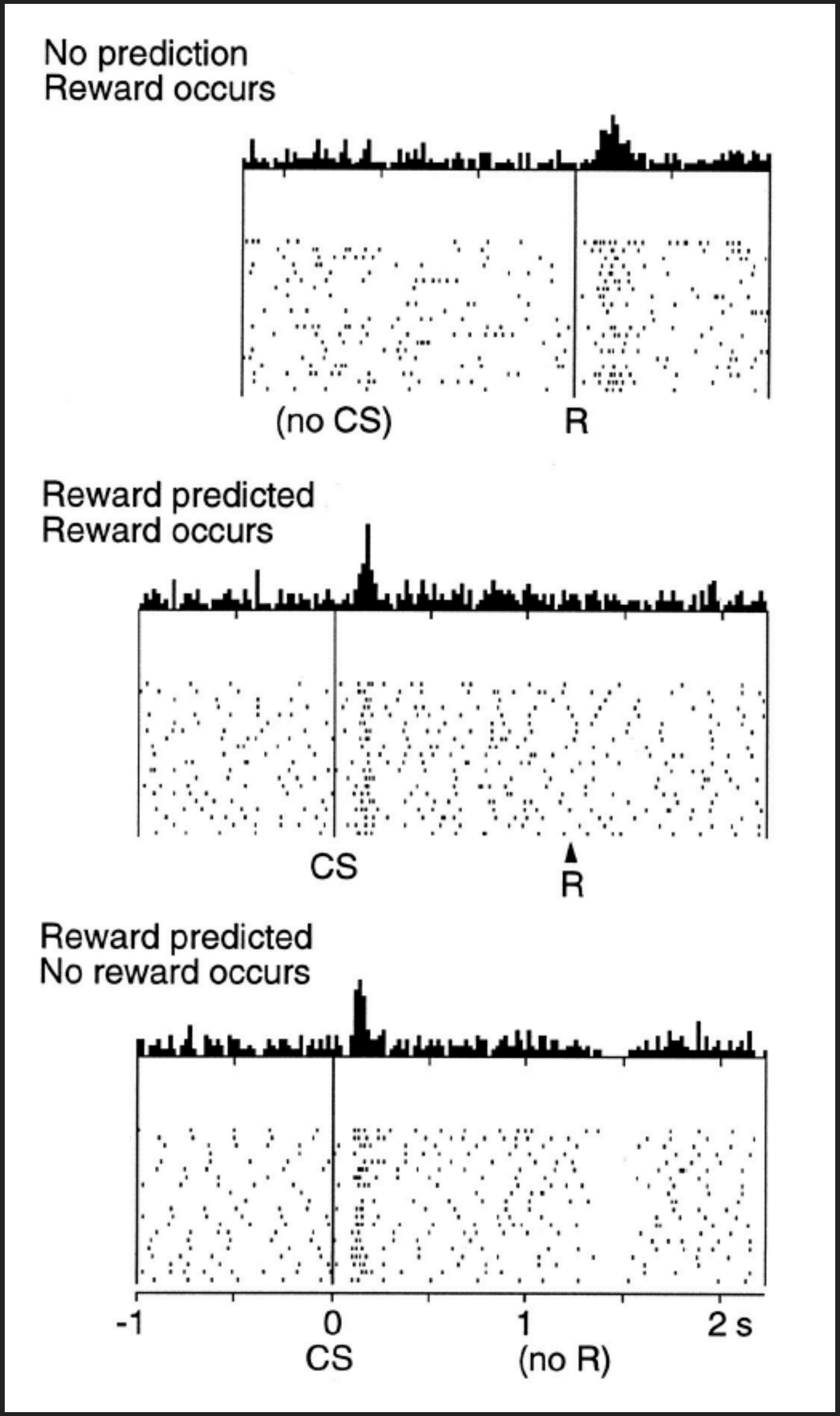
THE REWARD PREDICTION ERROR HYPOTHESIS

One of the functions of the phasic activity of dopamine-producing neurons in mammals is to deliver an error between an old and a new estimate of expected future reward to target areas throughout the brain.



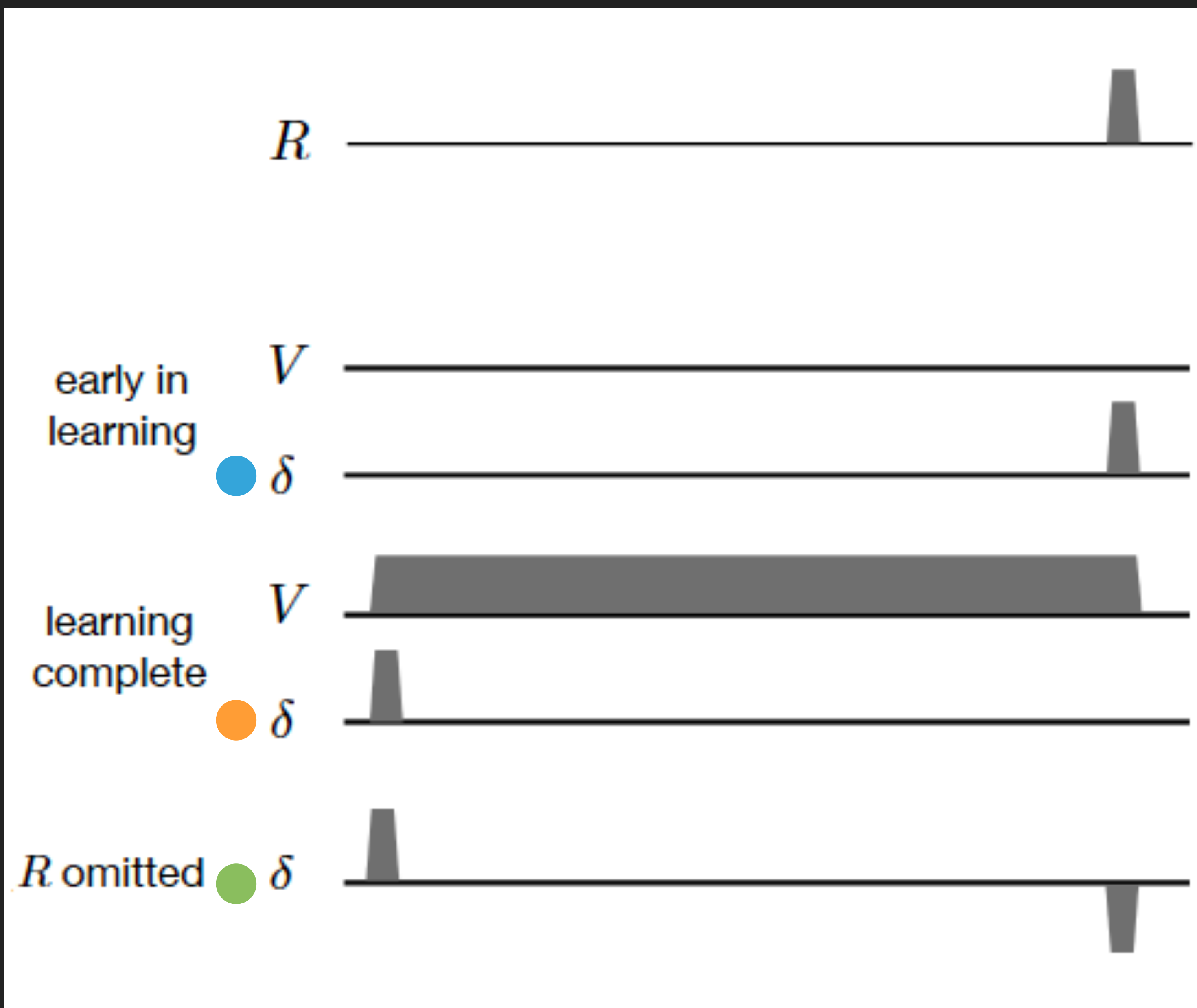
THE REWARD PREDICTION ERROR HYPOTHESIS

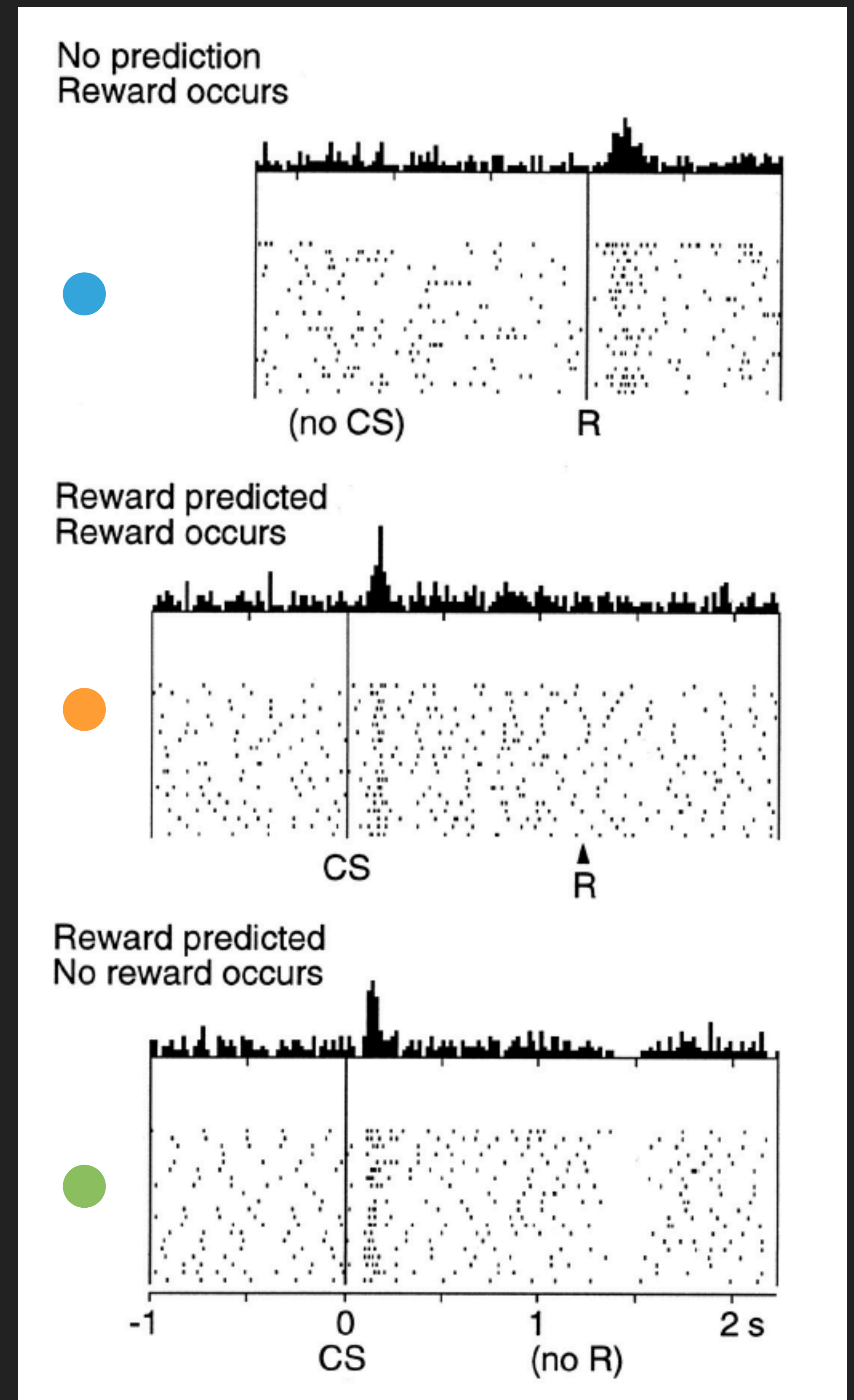
One of the functions of the phasic activity of dopamine-producing neurons in mammals is to deliver an error between an old and a new estimate of expected future reward to target areas throughout the brain.



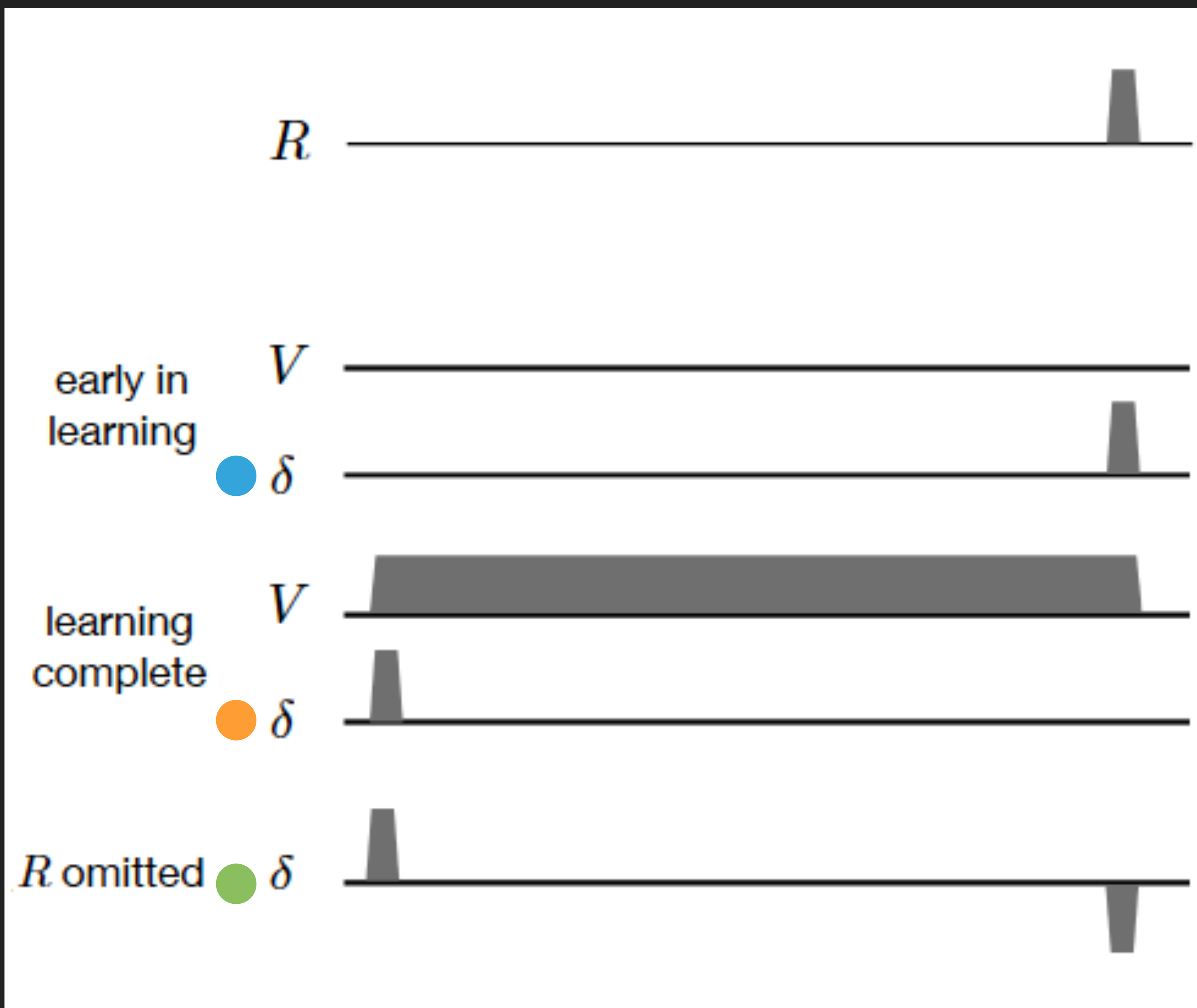
Schultz et al. (1997)

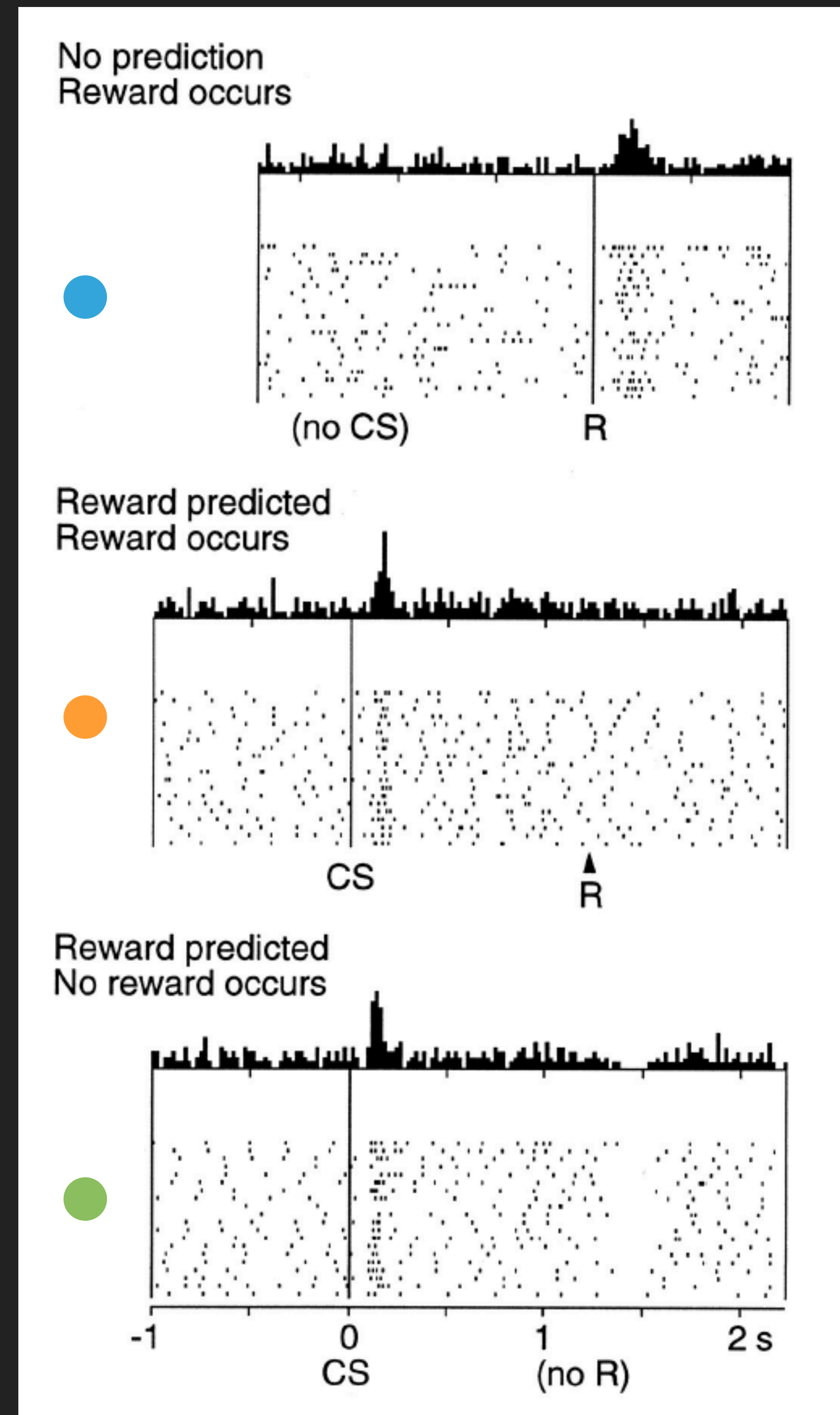
THE REWARD PREDICTION ERROR HYPOTHESIS





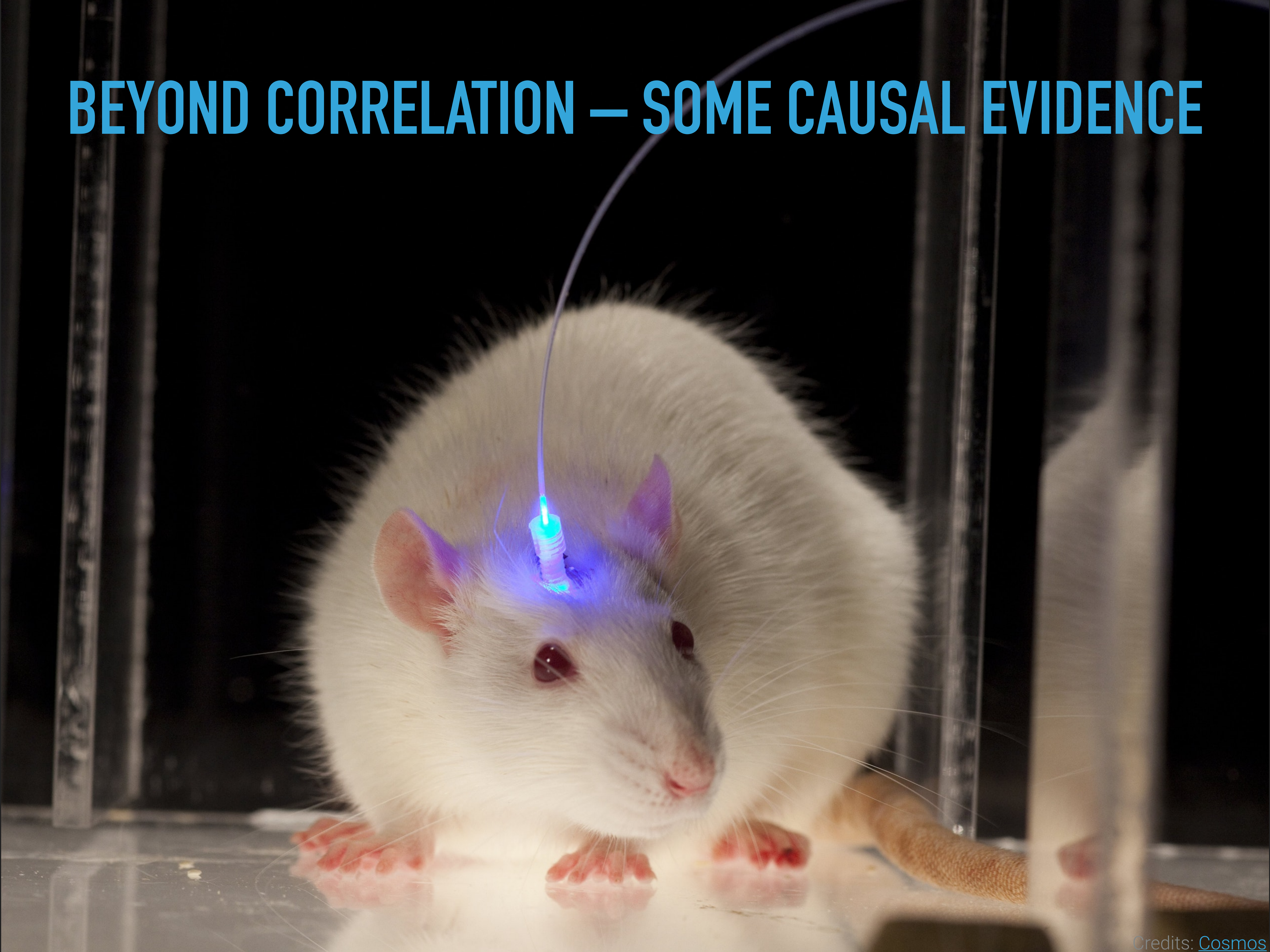
THE REWARD PREDICTION ERROR HYPOTHESIS





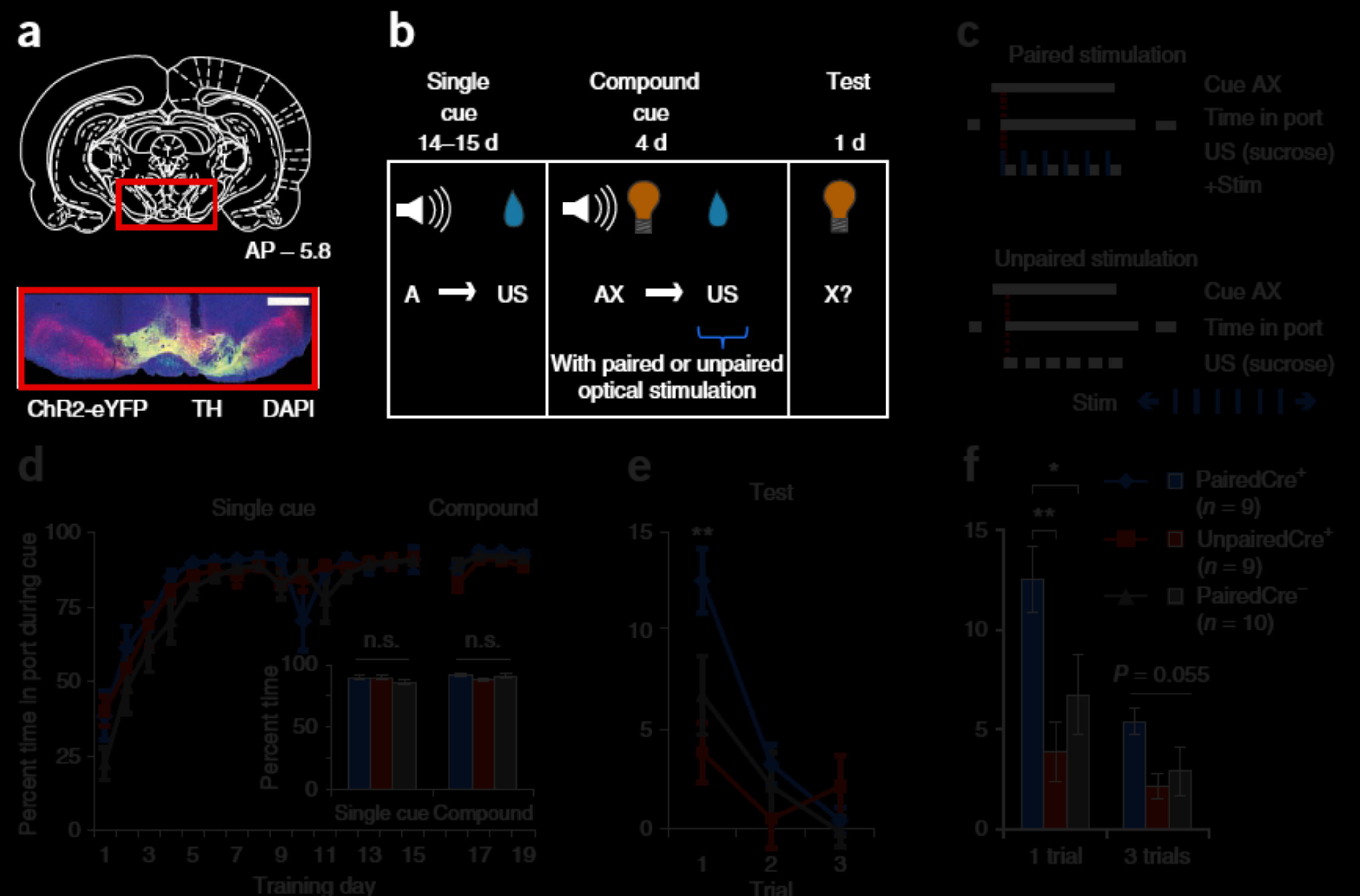
Schultz et al. (1997)

BEYOND CORRELATION – SOME CAUSAL EVIDENCE



Credits: [Cosmos]

BEYOND CORRELATION – SOME CAUSAL EVIDENCE



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

TAKEAWAY

Dopamine drives learning.

WHERE DO WE GO FROM HERE?

WHERE DO WE GO FROM HERE?

- What theories and results from Psychology and Neuroscience can help in building the computational model of the mind?

WHERE DO WE GO FROM HERE?

- What theories and results from Psychology and Neuroscience can help in building the computational model of the mind?
    - How does the brain construct a representation from the perception, before action? Is there even a separation?

WHERE DO WE GO FROM HERE?

- What theories and results from Psychology and Neuroscience can help in building the computational model of the mind?
    - How does the brain construct a representation from the perception, before action? Is there even a separation?
    - How does the brain encode, retrieve, and combine knowledge of different kinds, learned at different times? How does memory work?

WHERE DO WE GO FROM HERE?

- What theories and results from Psychology and Neuroscience can help in building the computational model of the mind?
  - How does the brain construct a representation from the perception, before action? Is there even a separation?
  - How does the brain encode, retrieve, and combine knowledge of different kinds, learned at different times? How does memory work?
  - How are high-level actions performed by combining low-level muscle-twitches? Is there some sort of hierarchical and modular arrangement of actions?

REFERENCES

- Kehoe, E. J., Schreurs, B. G., & Graham, P. (1987). Temporal primacy overrides prior training in serial compound conditioning of the rabbit's nictitating membrane response. *Animal Learning & Behavior*, *15*(4), 455-464.

- Ludvig, E. A., Bellemare, M. G., & Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications,* 111-144.

- Schultz, W., Dayan, P., Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1598.

- Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966.

- Sutton, R. S., Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 497-537.

- Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT press.