Reinforcement Learning in Psychology and Neuroscience

with thanks to
Elliot Ludvig
Princeton University

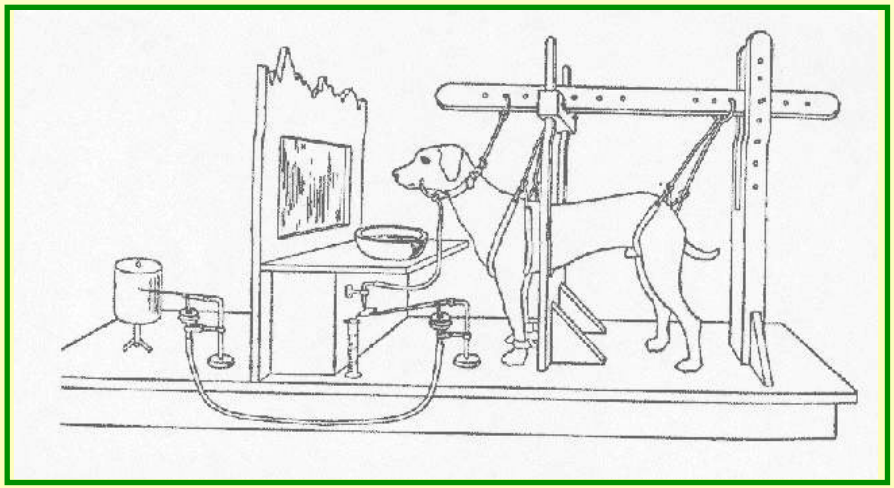
Psychology has identified two primitive kinds of learning

- *Classical* Conditioning
- *Operant* Conditioning (a.k.a. Instrumental learning)
- Computational theory:
  - *Classical* = Prediction
    - What is going to happen?
  - *Operant* = Control
    - What to do to maximize reward?

Classical Conditioning

Pavlov

- Russian physiologist
- Interested in how learning happened in the brain
- Conditional and Unconditional Stimuli



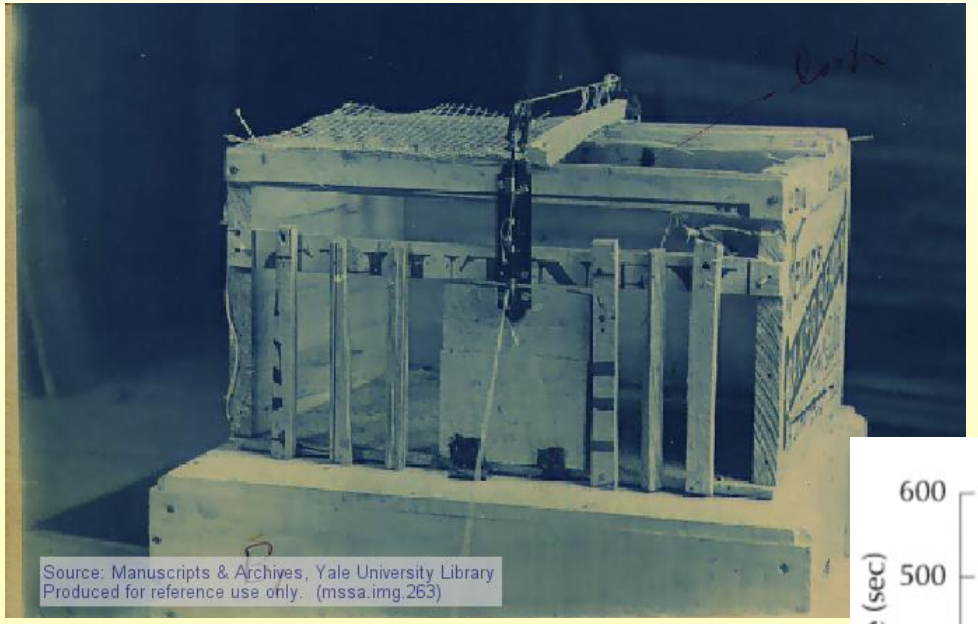
Rescorla-Wagner Model (1972)

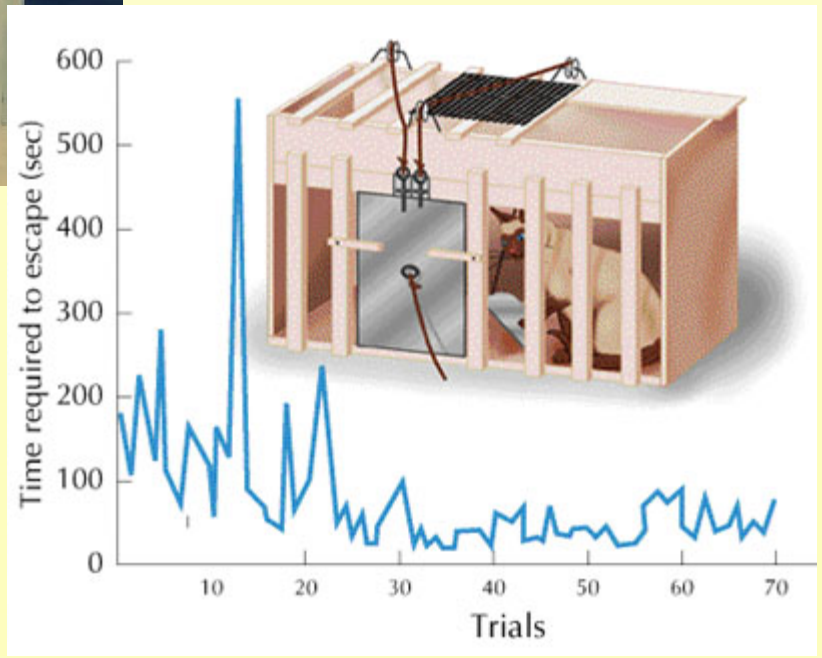
- Computational model of conditioning
  - Widely cited and used
- Learning as violation of expectations
  - TD learning as extension of RW

Operant Learning

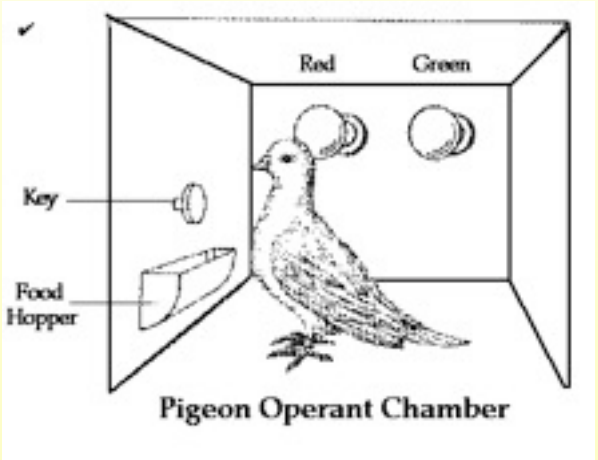
- Operant Conditioning is all about choice in 3 main ways:
  - Decide which response to make?
  - Decide how much to respond?
  - Decide when to respond?

Thorndike's Puzzle Box





Operant Chambers



Pigeon Operant Chamber

Complex Cognition

Marr's 3 Levels of Analysis

- Computational
  - What function is being fulfilled?
- Algorithmic
  - How is it accomplished?
- Implementational
  - What physical substrate is involved?

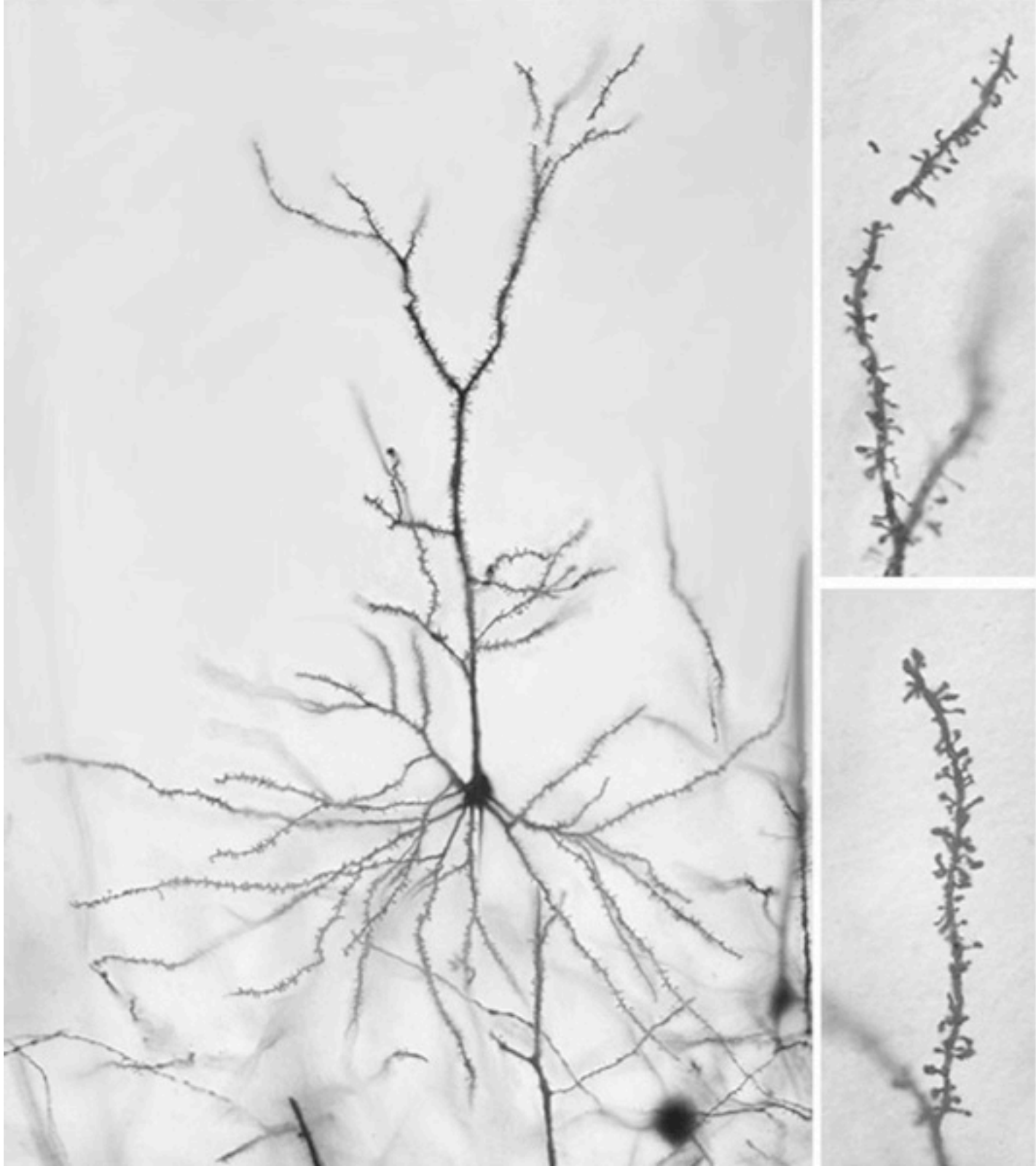
The Basic TD Model

- Learn to predict discounted sum of upcoming reward through TD with linear function approximation:

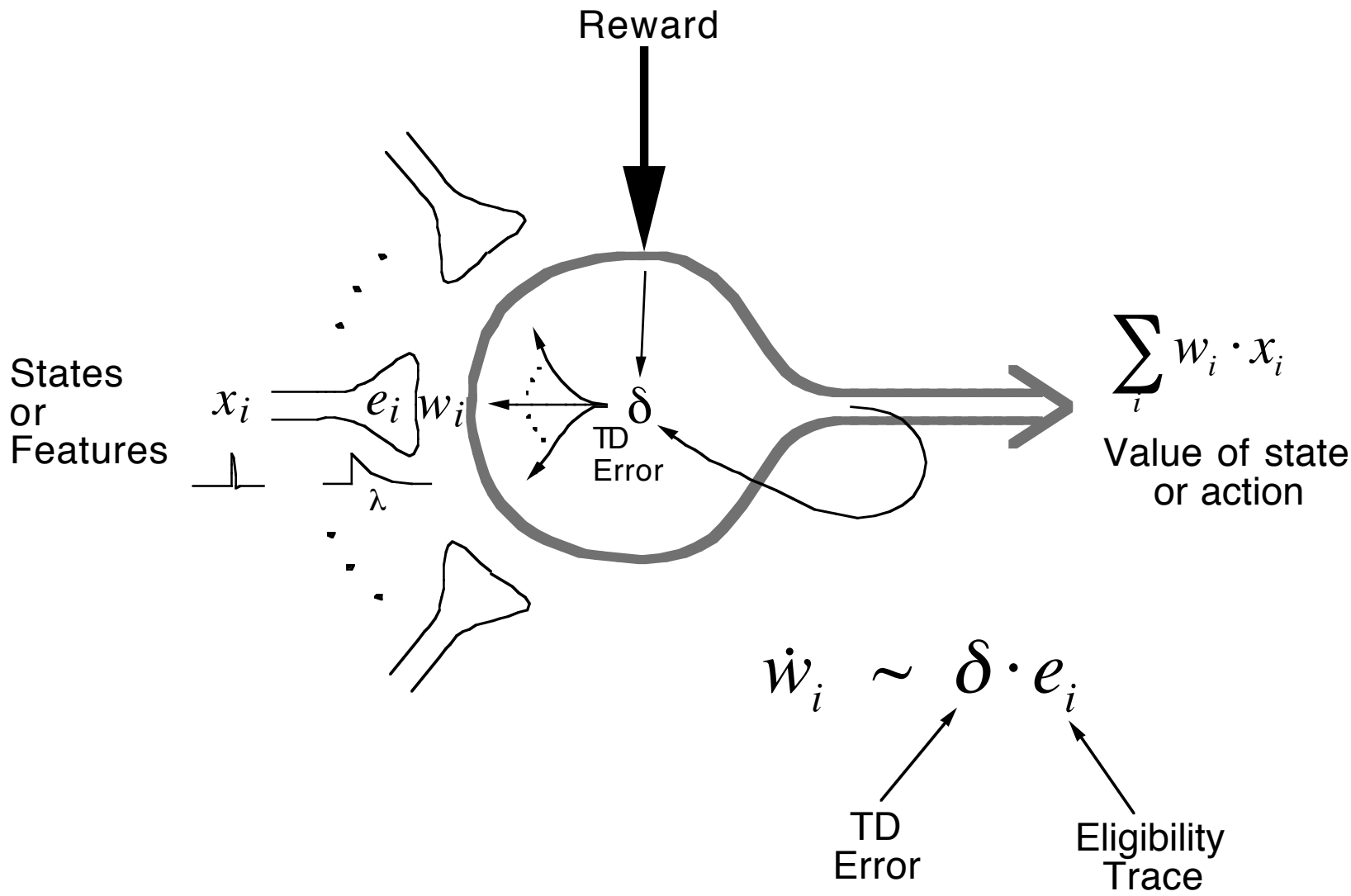
$$V_t = \mathbf{w}_t^T \mathbf{x}_t = \sum_{i=1}^{n} w_t(i) x_t(i)$$

- The TD error is calculated as:

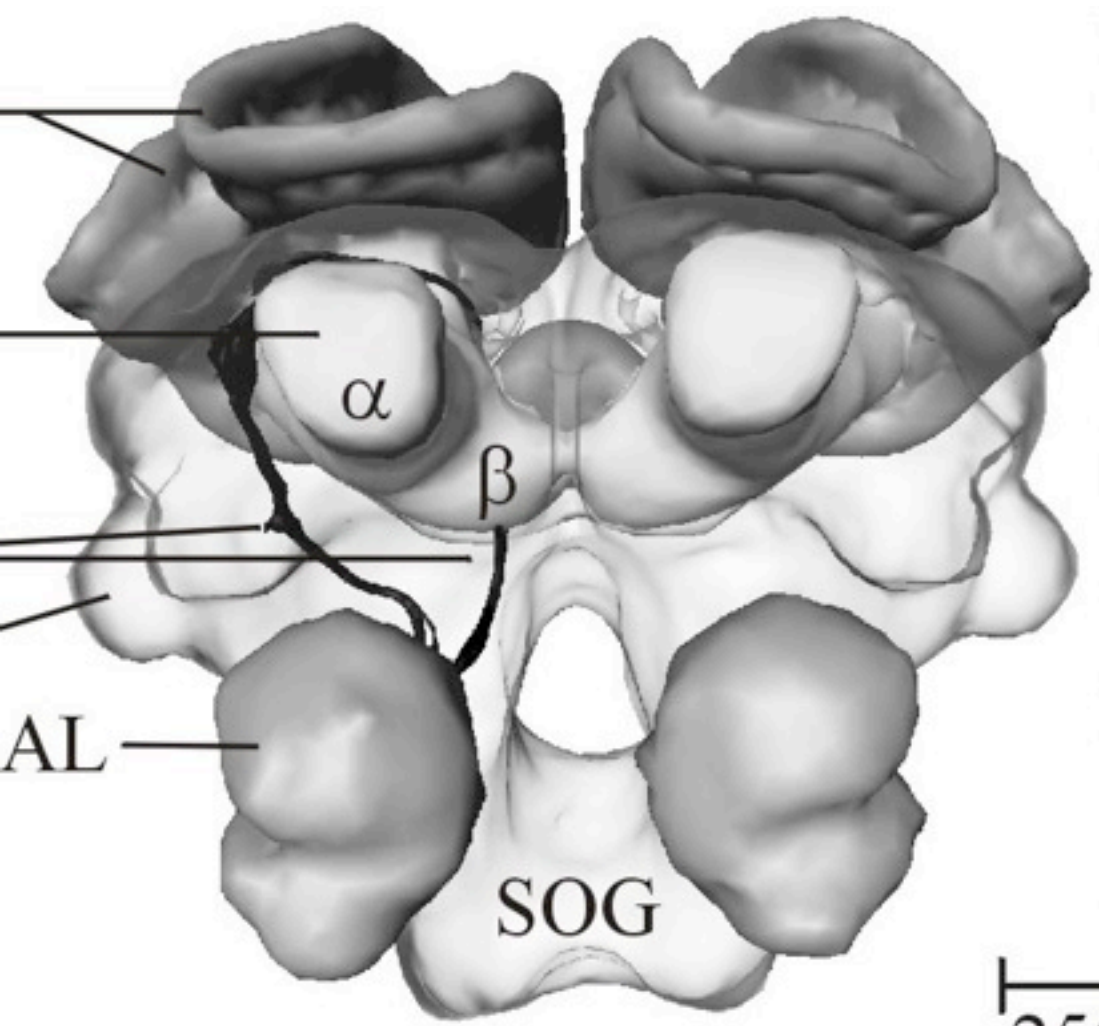
$$\delta_t = r_{t+1} + \gamma V_{t+1} - V_t$$



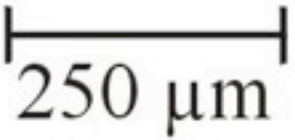
TD(λ) algorithm/model/neuron

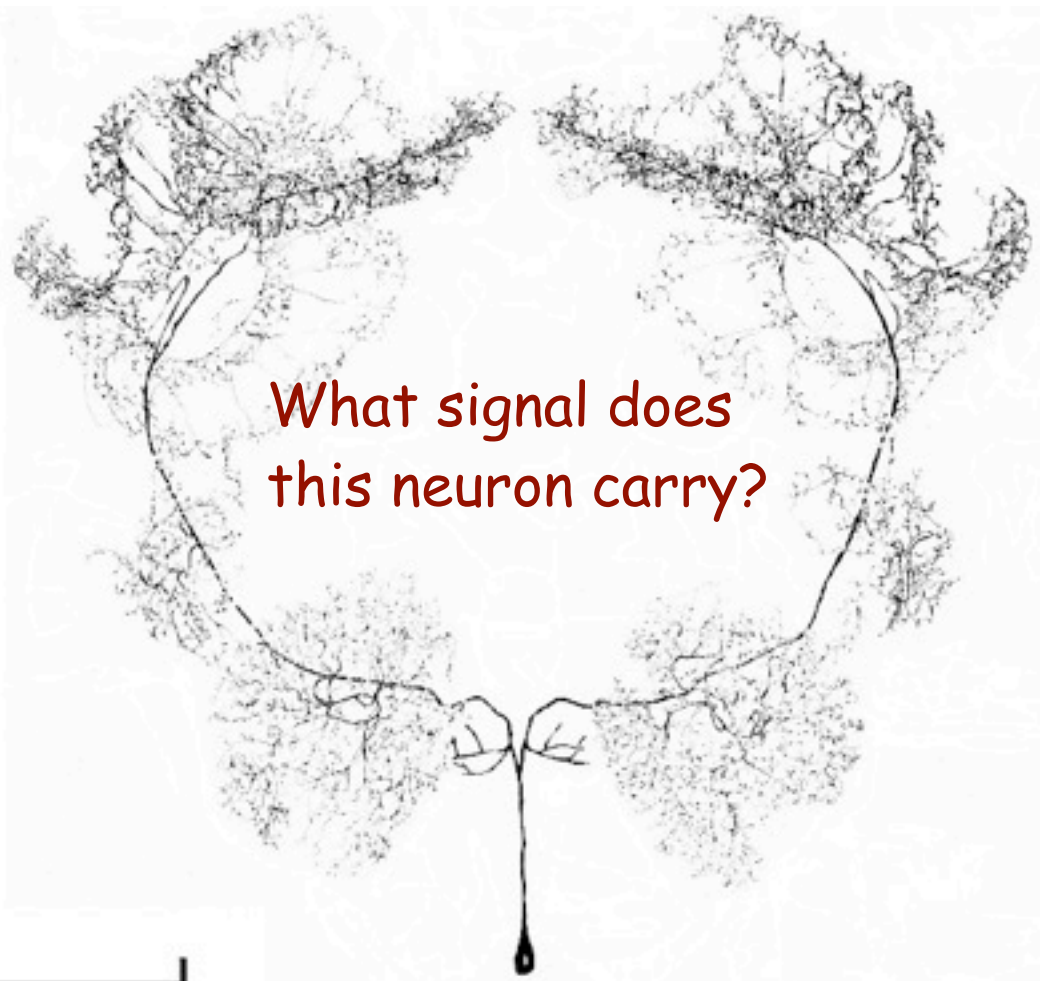


Brain reward systems



Honeybee Brain





VUM Neuron

Hammer, Menzel

Dopamine

- Small-molecule Neurotransmitter
  - Diffuse projections from mid-brain throughout the brain

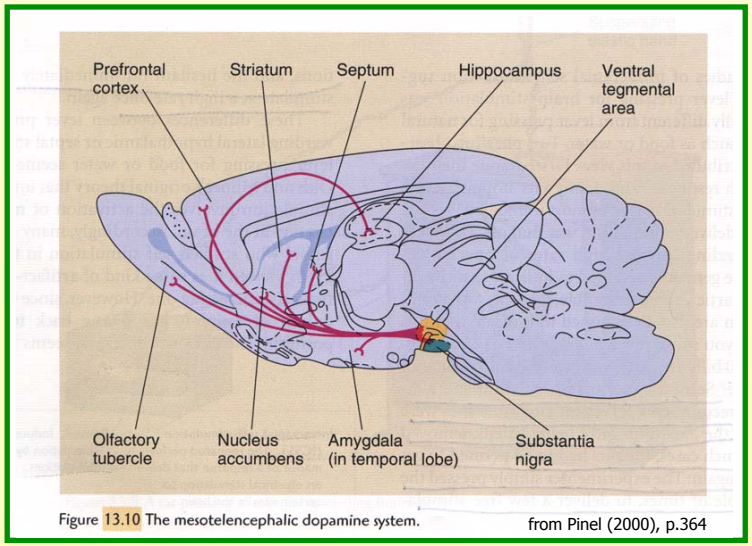
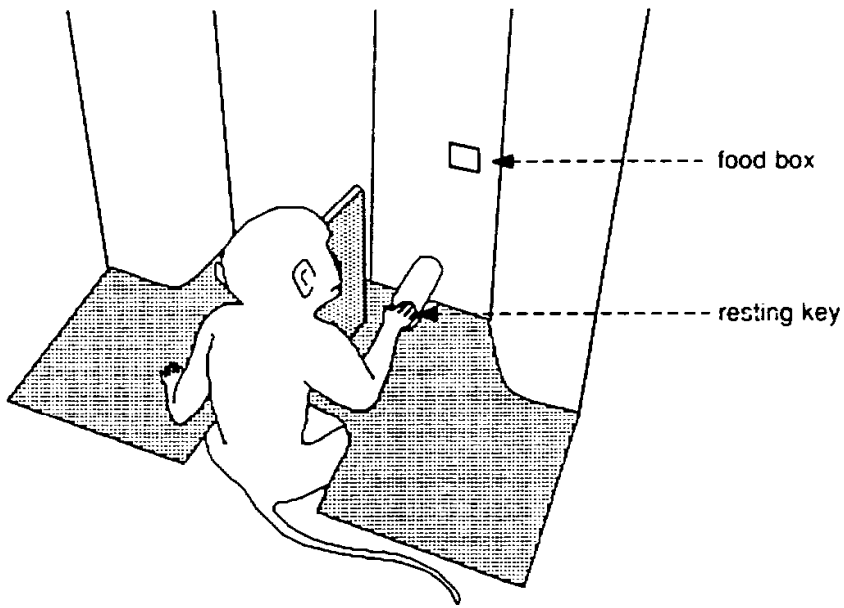


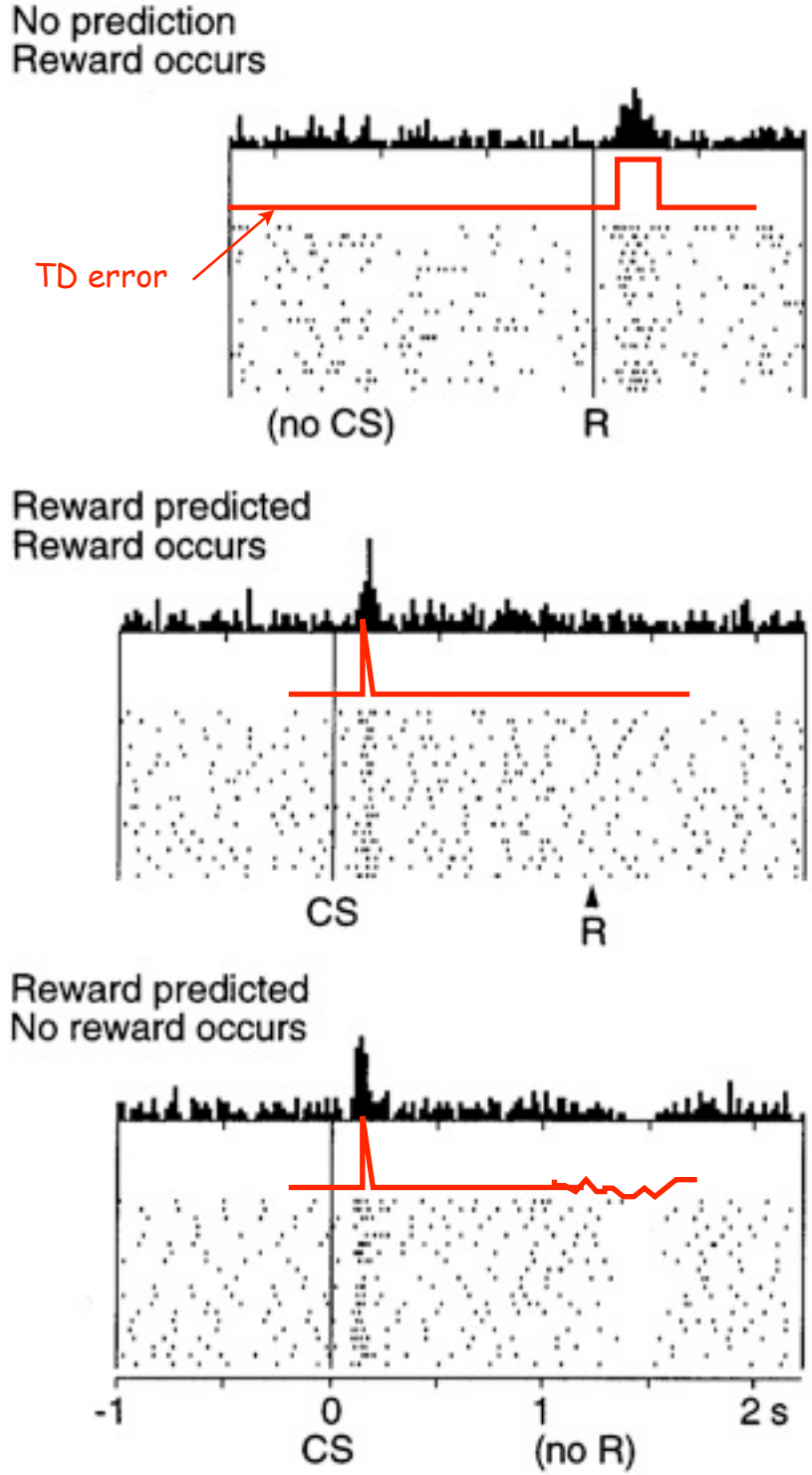
Figure 13.10 The mesotelencephalic dopamine system. from Pinel (2000), p.364

Key Idea: Phasic change in baseline dopamine responding = reward prediction error

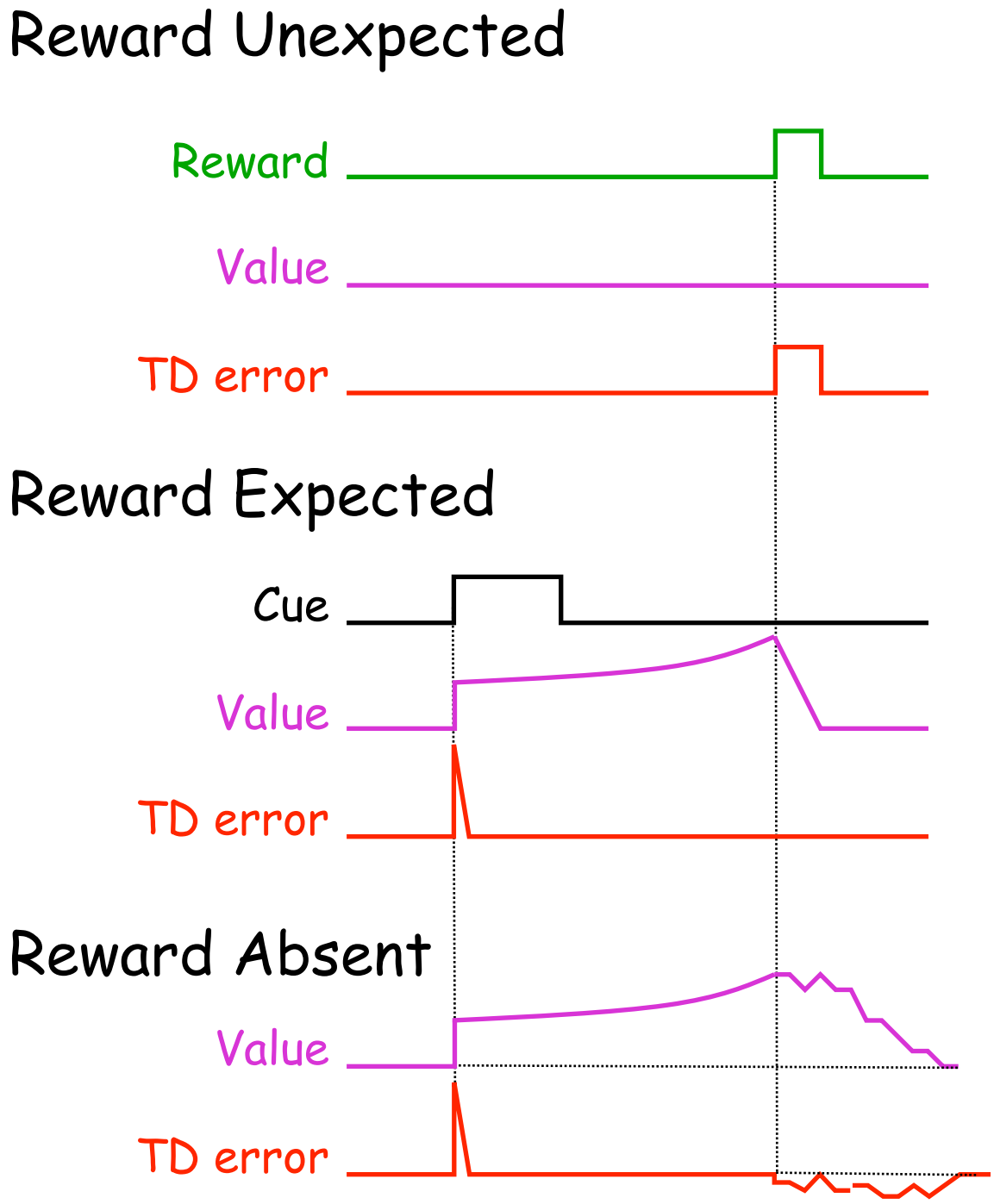
Dopamine neurons signal the error/change in prediction of reward



Wolfram Schultz, et al.



Representation-independent predictions of TD errors



$$TD\ error_t = r_{t+1} + \gamma V_{t+1} - V_t$$

The theory that *Dopamine = TD error*
is one of the *most important interactions ever*
between artificial intelligence and neuroscience