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?
Operant Chambers
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:

\[ V_t = w^T_t x_t = \sum_{i=1}^{n} w_t(i) x_t(i) \]

- The TD error is calculated as:

\[ \delta_t = r_{t+1} + \gamma V_{t+1} - V_t \]
TD(λ) algorithm/model/neuron

\[ \dot{w}_i \sim \delta \cdot e_i \]

Reward

Value of state or action

\[ \sum_i w_i \cdot x_i \]

TD Error

Eligibility Trace

States or Features

\[ x_i, e_i, w_i \]

\[ TD(\lambda) \text{ error} \]

\[ \delta \cdot e_i \]

\[ \sum_i w_i \cdot x_i \]
Brain reward systems

What signal does this neuron carry?

Honeybee Brain

VUM Neuron

Hammer, Menzel
Dopamine

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

Key Idea: Phasic change in baseline dopamine responding = reward prediction error
Dopamine neurons signal the error/change in prediction of reward
Representation-independent predictions of TD errors

\[ \text{TD error}_t = r_{t+1} + \gamma V_{t+1} - V_t \]
The theory that \textit{Dopamine = TD error} is one of the \textit{most important interactions ever} between artificial intelligence and neuroscience.