

Reinforcement Learning and Psychology: A Personal Story

Rich Sutton
Department of Computing Science
University of Alberta

Personal perspective

- There is a *science of mind* that is neither natural science nor applications technology
- In the future, most minds will be *designed* rather than evolved
- Reinforcement learning is the beginning of an interdisciplinary, multi-level science of mind
- The origins and heart of reinforcement learning are in psychology

Marr's Three Levels

at which any information processing system
can be understood

- **Computational Theory Level**

- What are the goals of the computation? What and Why?
- What is being computed?
- Why are these the right things to compute?

- **Representation and Algorithm Level**

- How are these things computed? How?
- What representation and algorithms are used?

- **Hardware Implementation Level**

- How is this implemented physically? Really how?

The *most important interaction ever* between psychology and the engineering sciences may be the theory that brain reward systems are implementing reinforcement learning algorithms

in particular, that *Dopamine = TD error*



Martin Hammer
~1995



Wolfram Schultz



Read Montague

Science, 1997



Peter Dayan

Outline

1. The “discovery” of reinforcement learning
 - that instrumental learning was missing from the engineering sciences
2. The discovery of temporal-difference learning
 - in classical conditioning, as engineering, and in brain reward systems
3. (Planning as RL on imagined experience)

Reinforcement learning

- The engineering endeavor most closely related to natural learning in animals and people
- A new (~30 year old) class of learning algorithms, inspired by animal learning psychology, and developed within machine learning and AI, for approximately solving large optimal-control problems
- RL methods have outperformed previous solution methods in many cases: Game-playing, robot control, auto-pilots, efficient management of queues, inventories, power systems...
- RL ideas provide a computational theory that deepens our understanding of natural learning behavior and mechanisms

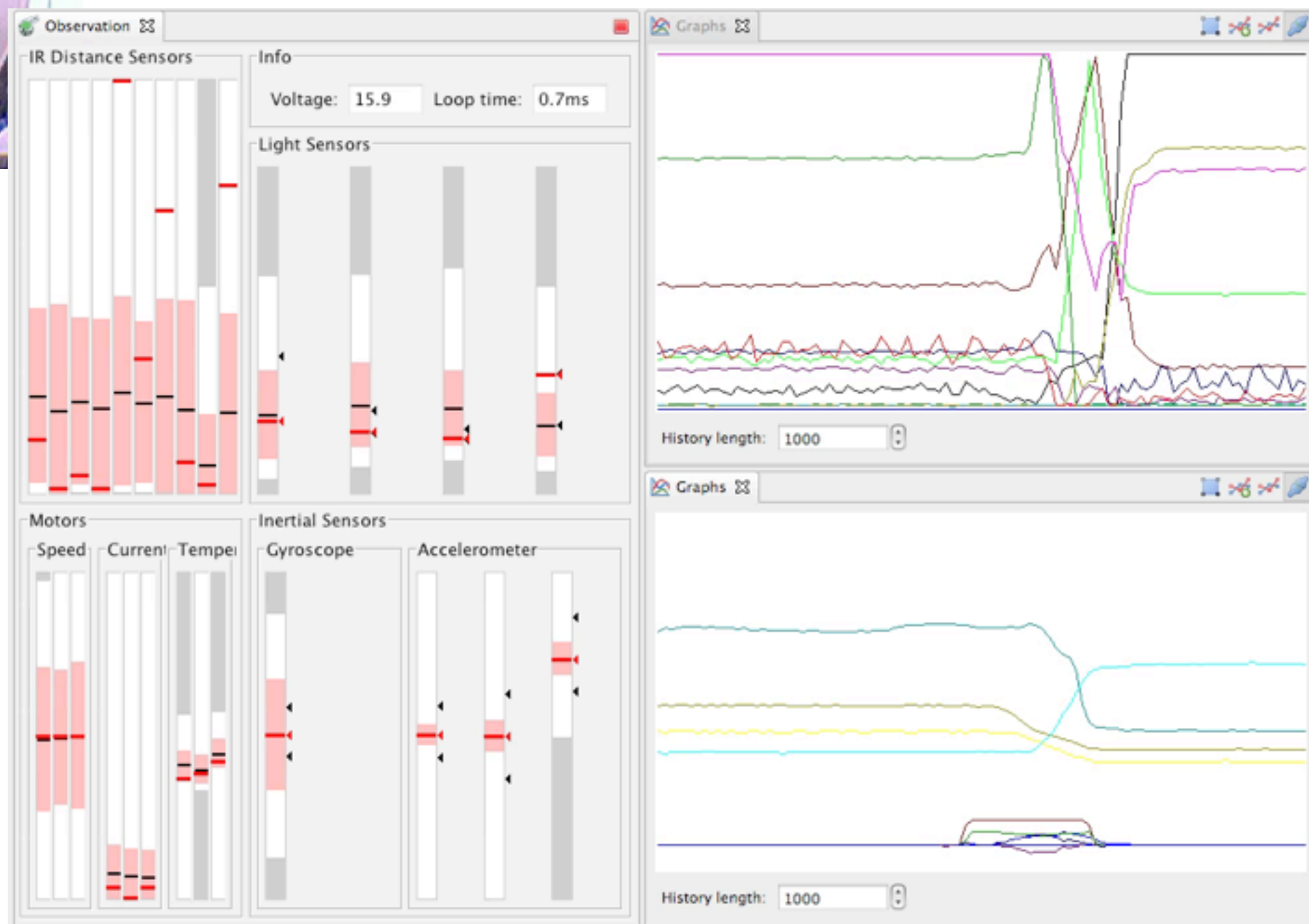
Real-time learning in the critterbot





Citterbot wall-following behavior

Citterbot signal data

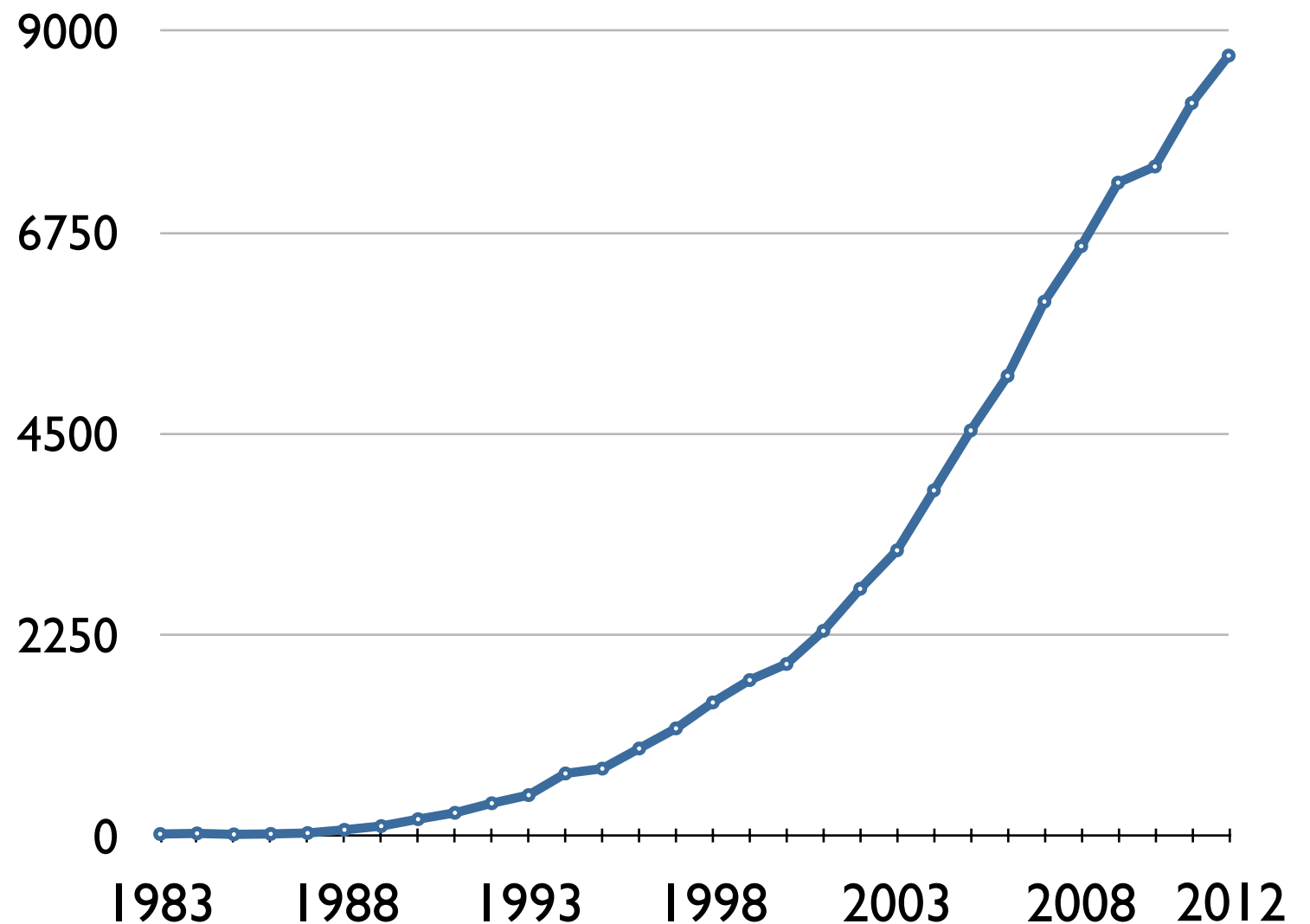


Reinforcement learning

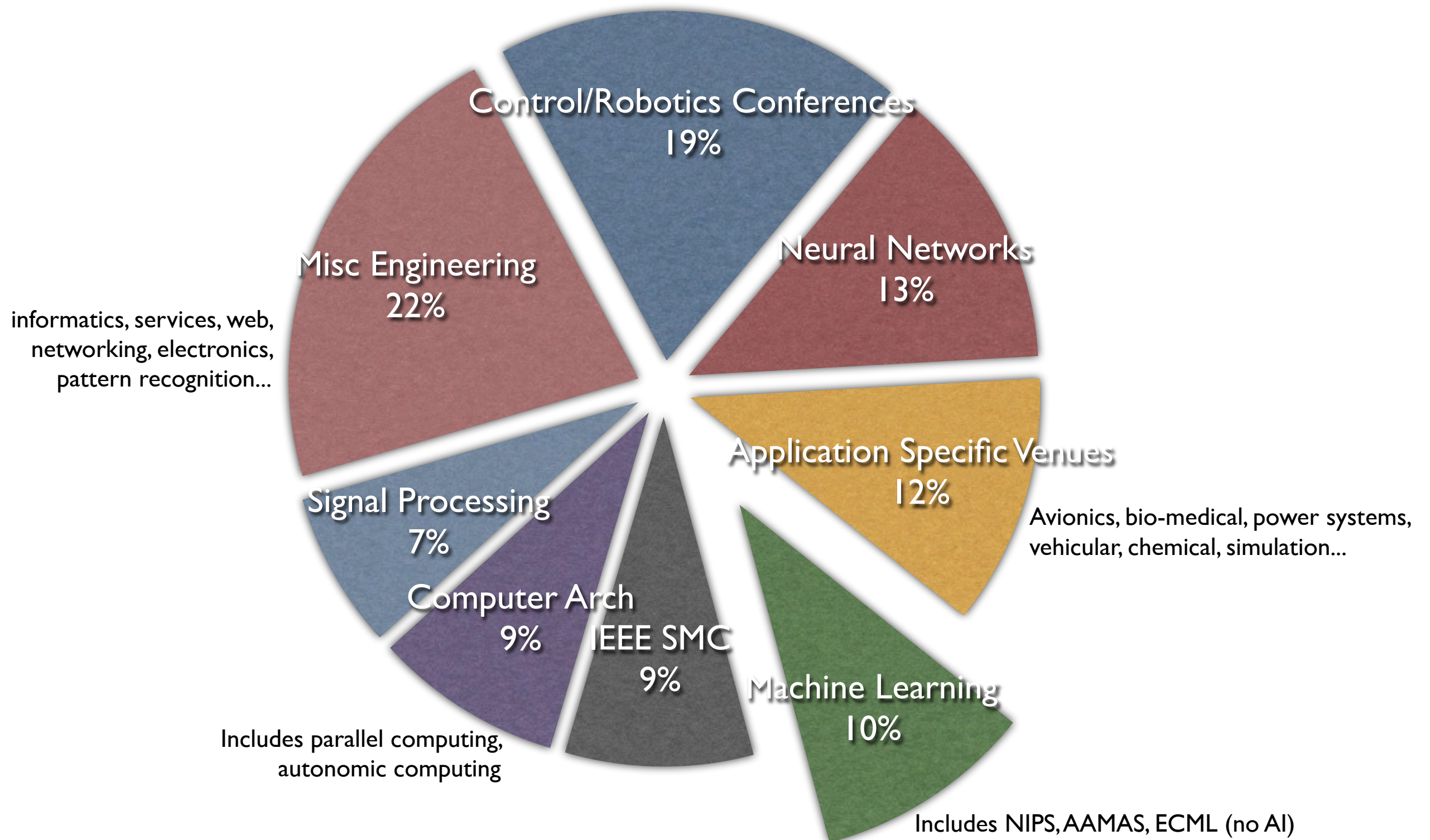
- The engineering endeavor most closely related to natural learning in animals and people
- A new (~30 year old) class of learning algorithms, inspired by animal learning psychology, and developed within machine learning and AI, for approximately solving large optimal-control problems
- RL methods have outperformed previous solution methods in many cases: Game-playing, robot control, auto-pilots, efficient management of queues, inventories, power systems...
- RL ideas provide a computational theory that deepens our understanding of natural learning behavior and mechanisms

RL has grown rapidly

Google scholar
hits/year for
“reinforcement
learning”



Fields publishing RL applications



Reinforcement learning

- The engineering endeavor most closely related to natural learning in animals and people
- A new (~30 year old) class of learning algorithms, inspired by animal learning psychology, and developed within machine learning and AI, for approximately solving large optimal-control problems
- RL methods have outperformed previous solution methods in many cases: Game-playing, robot control, auto-pilots, efficient management of queues, inventories, power systems...
- RL ideas provide a computational theory that deepens our understanding of natural learning behavior and mechanisms




Stanford University Autonomous Helicopter

Andrew Ng, Pieter Abbeel, Adam Coates, et al.

Reinforcement learning

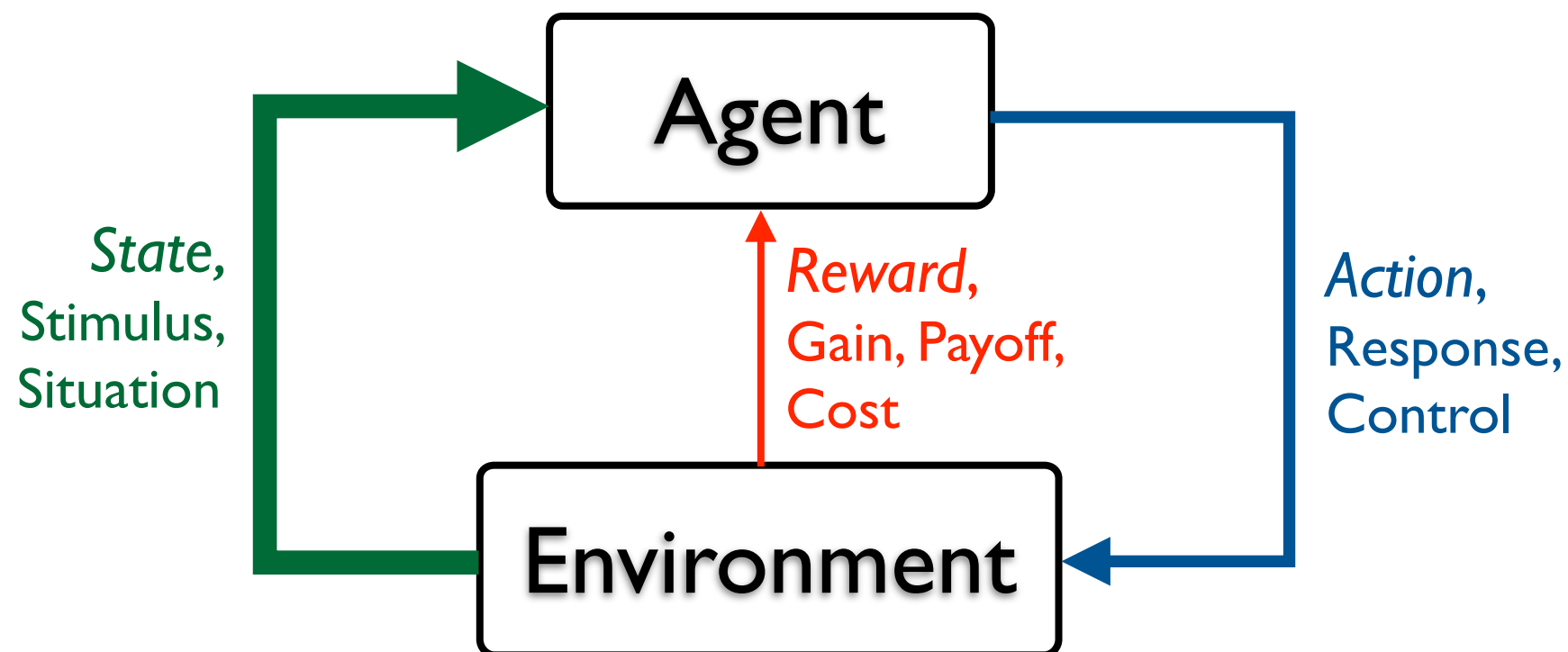
- The engineering endeavor most closely related to natural learning in animals and people
- A new (~30 year old) class of learning algorithms, inspired by animal learning psychology, and developed within machine learning and AI, for approximately solving large optimal-control problems
- RL methods have outperformed previous solution methods in many cases: Game-playing, robot control, auto-pilots, efficient management of queues, inventories, power systems...
- RL ideas provide a computational theory that deepens our understanding of natural learning behavior and mechanisms

Outline

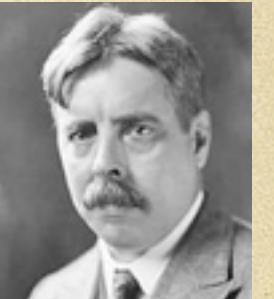
- 
1. The “discovery” of reinforcement learning
 - that instrumental learning was missing from the engineering sciences
 2. The discovery of temporal-difference learning
 - in classical conditioning, as engineering, and in brain reward systems
 3. (Planning as RL on imagined experience)

The “discovery” of reinforcement learning

- of learning by trial and error how to act so as to maximize a received scalar signal



Thorndike's "Law of Effect"



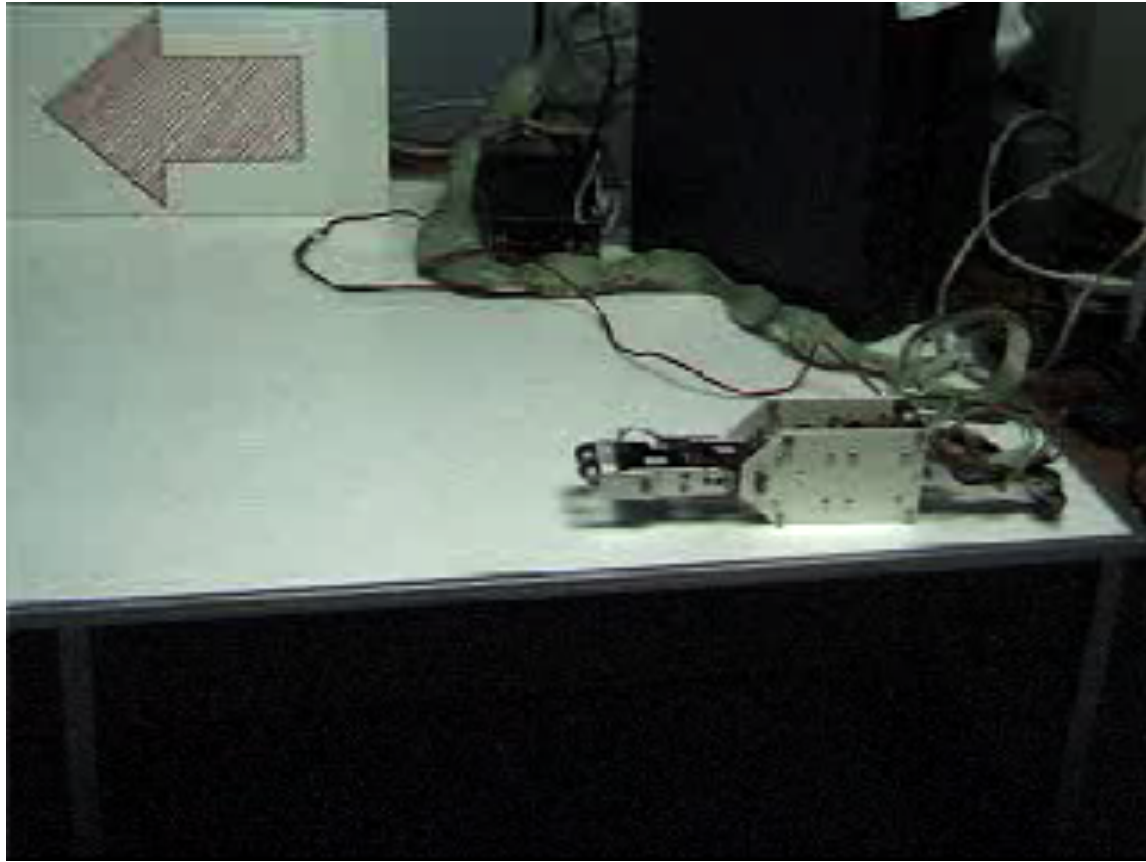
"Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur.

—Thorndike, 1911

"If it feels good, do it."

—Anonymous

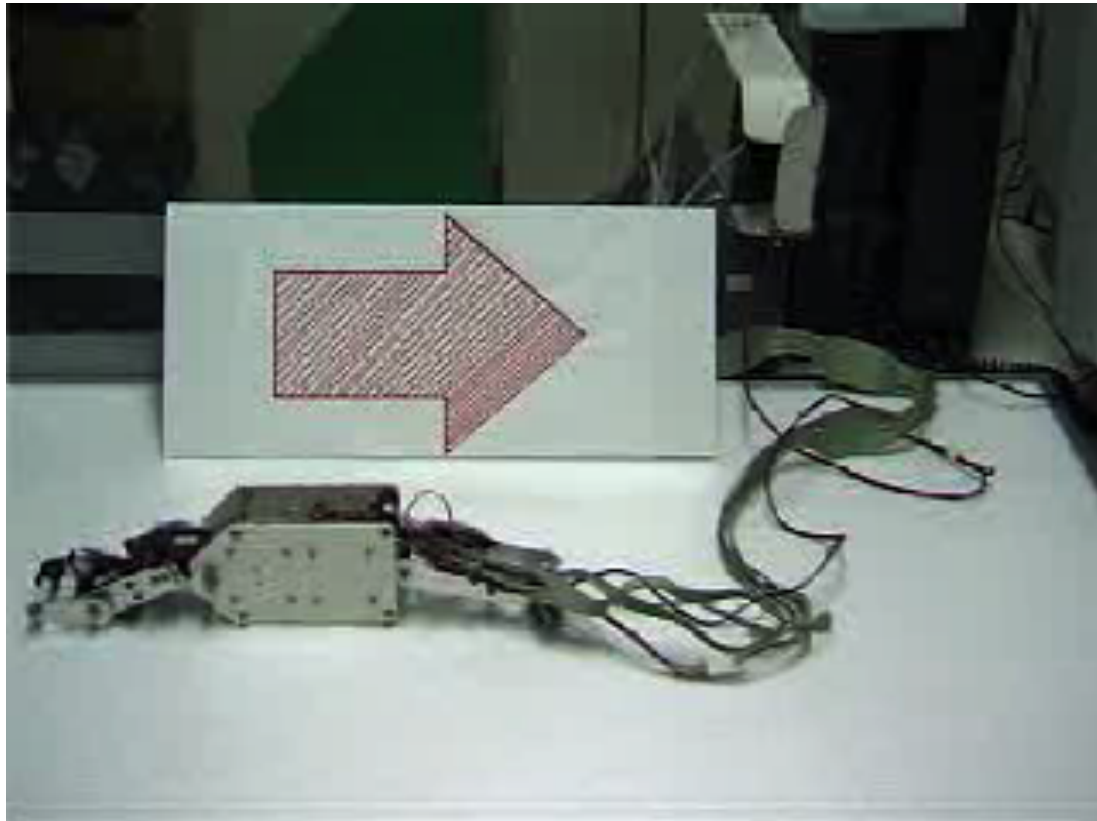
Hajime Kimura's RL Robots



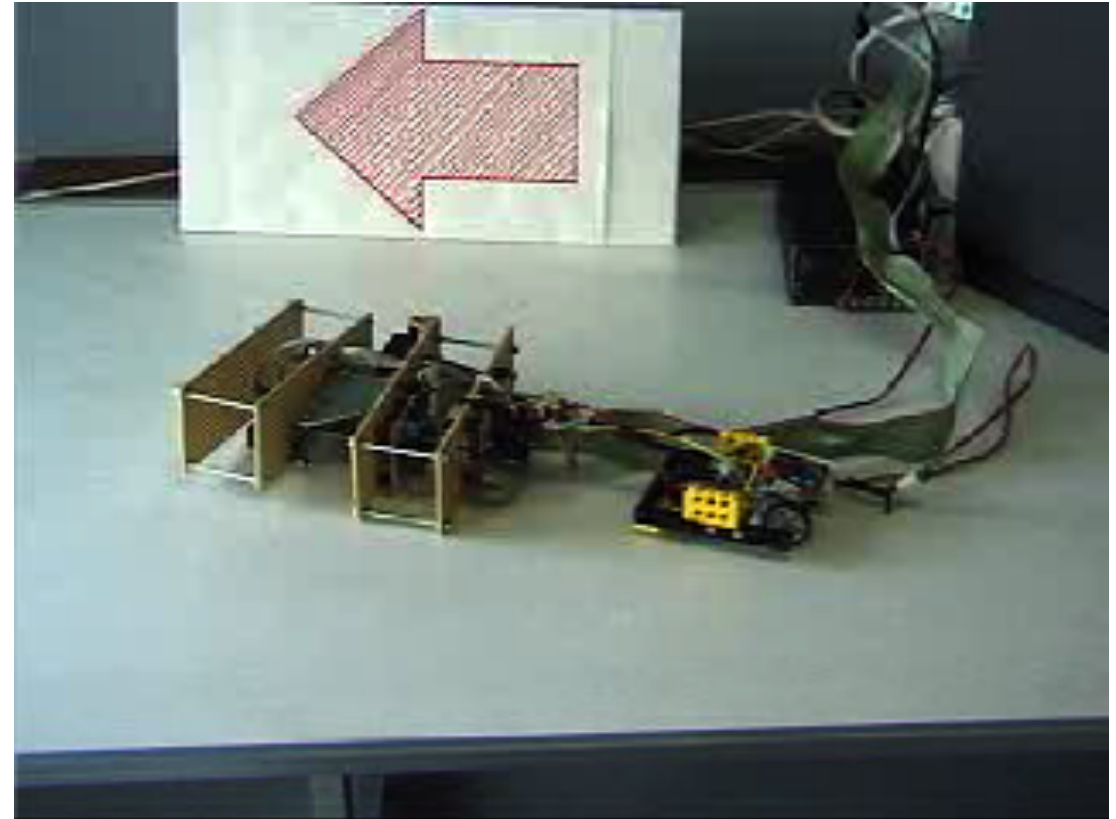
Before



After



Backward



New Robot, Same algorithm

The reward hypothesis

“That all of what we mean by goals and purposes can be well thought of as maximization of the expected value of the cumulative sum of a received scalar signal (reward)”

—2004

Thus, this is an important problem to study from an engineering / computational point of view

Trial-and-error learning was missing from engineering/AI

- In psychology, there are clearly two kinds of associative learning phenomena: classical and instrumental
- Instrumental learning involves trial and error
 - behavior *affects* the animal's input
 - there is a need for *spontaneous variation* in behavior, to *search* for the optimal action-selection policy
- But in engineering, there was no clear analog of instrumental learning
- It took us—and the field—forever to realize this!



Harry Klopf
1941–1997



Andy Barto

Sometimes the obvious things are the hardest to see

- The discovery of gravity, by Isaac Newton
- The discovery of air/vacuum
- The discovery that people are animals, by Darwin et al.
- The discovery of reinforcement learning, by Harry Klopff, in the 1970s



Harry Klopff
1941–1997

Outline

1. The “discovery” of reinforcement learning

- that instrumental learning was missing from the engineering sciences

2. The discovery of temporal-difference learning

- in classical conditioning, as engineering, and in brain reward systems

3. (Planning as RL on imagined experience)

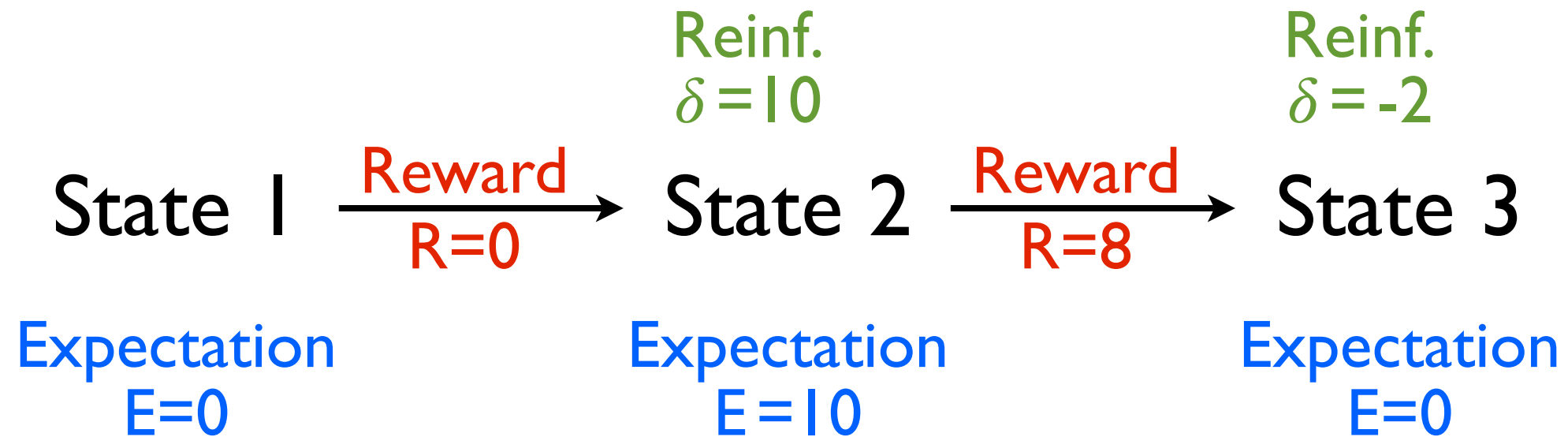
The mystery of expectation and reinforcement

- An expectation is a positive—as in a predictor of reward can act as a reward itself (secondary reinforcement)
- But an expectation is also a negative—actual reward has to be greater than expected in order to reinforce
- How can expectations contribute both positively and negatively to reinforcement?

The mystery of expectation and reinforcement

- An expectation is a positive—as in a predictor of reward can act as a reward itself (secondary reinforcement)
- But an expectation is also a negative—actual reward has to be greater than expected in order to reinforce
- How can expectations contribute both positively and negatively to reinforcement?
- only over time

Expectation and reinforcement—in real time



$$\delta_{t+1} = R_{t+1} + .9E_{t+1} - E_t$$

δ = Reinf. = The temporal-difference (TD) error
= reward-prediction error

The TD model of classical conditioning

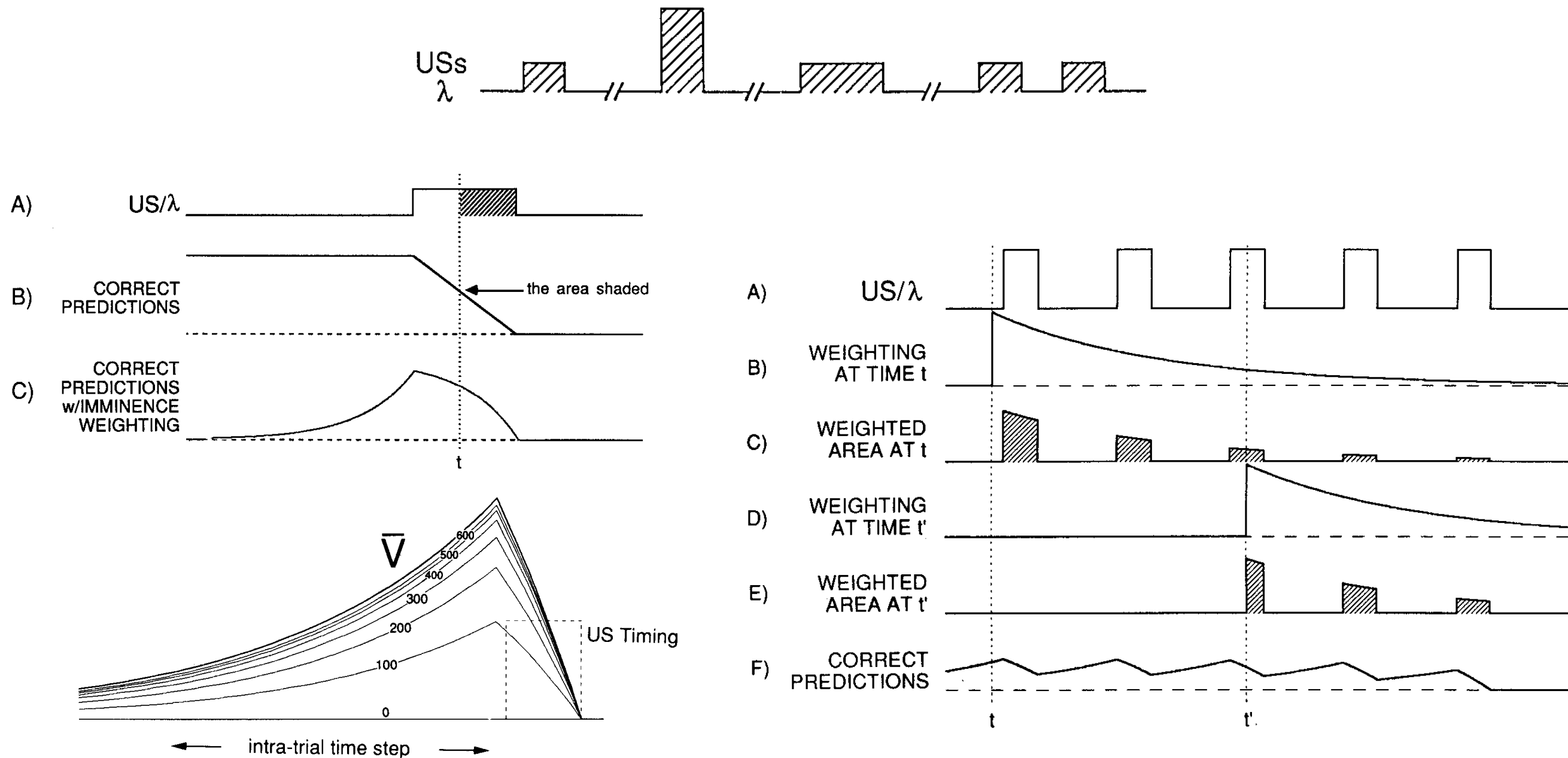
Expectations are sums of weights (associative strengths) $w(i)$, for each feature (CS component) i that is present at the time:

$$E_t = \sum_i w_t(i) \quad \text{for all features } i \text{ present at time } t$$

The weights of the present features are incremented in proportion to the TD error at the time:

$$\Delta w_t(i) \propto \delta_t \quad \text{for all features } i \text{ recently present at } t$$

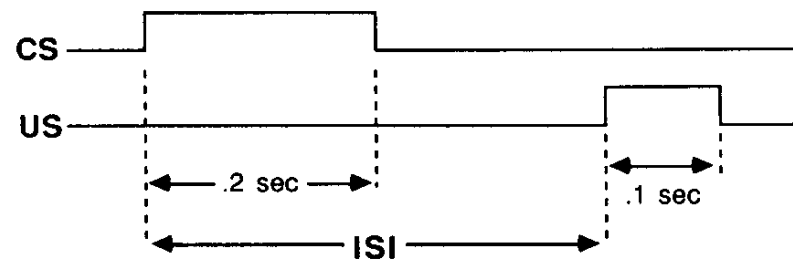
Q: What *real-time quantity* is learned?



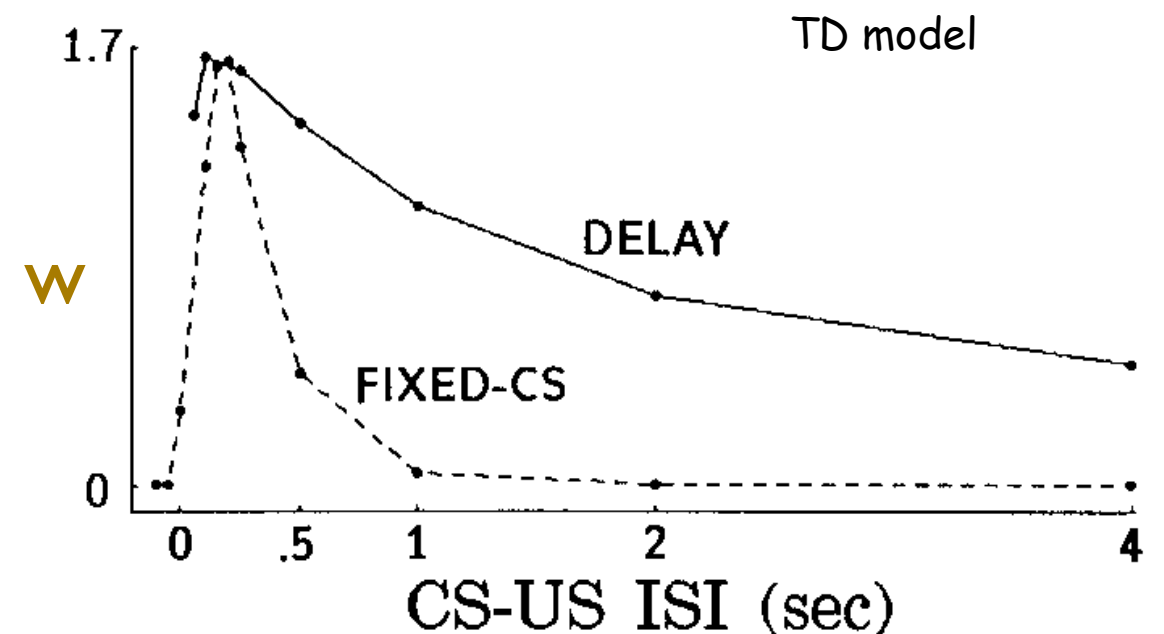
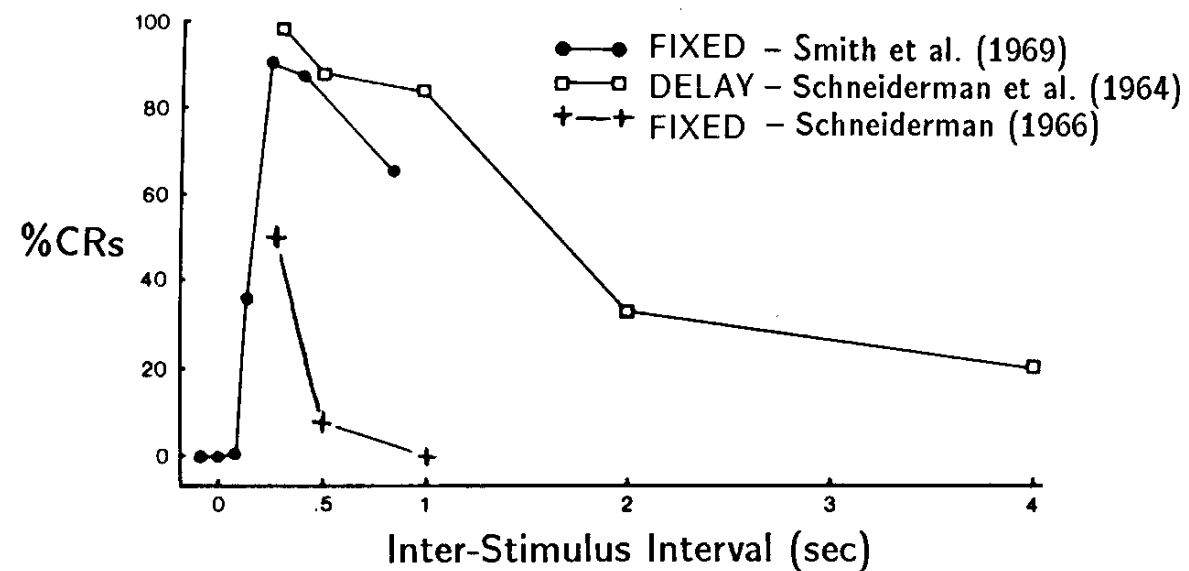
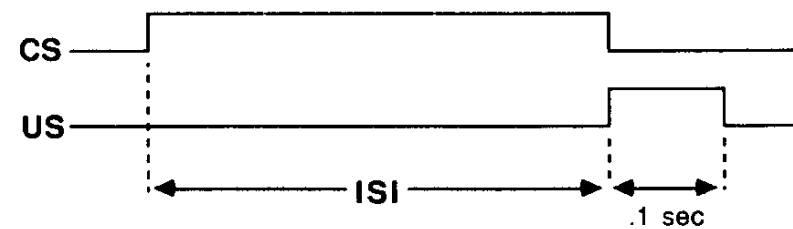
A: Expected discounted future reward

Effect of inter-stimulus interval

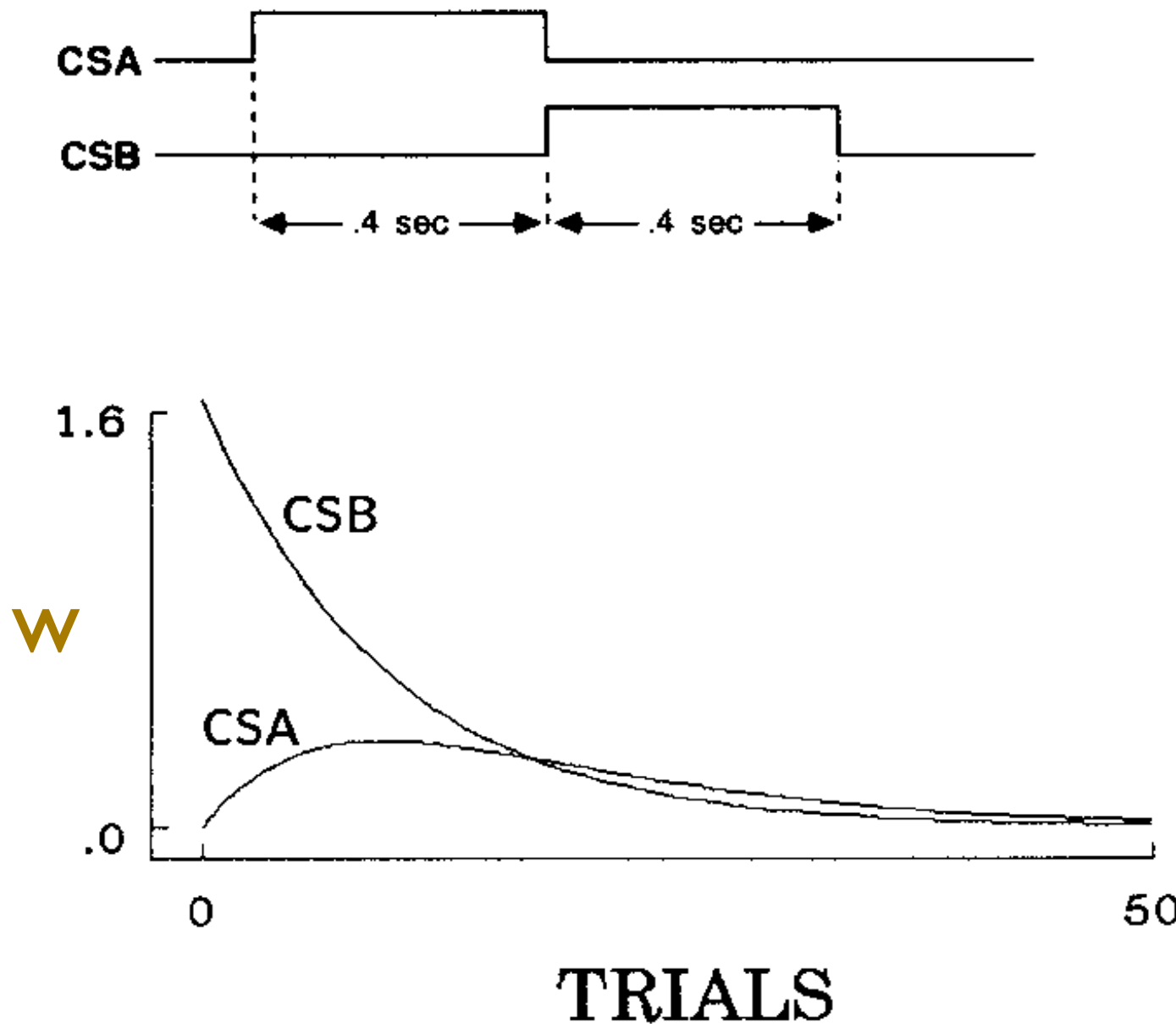
FIXED-CS CONDITIONING



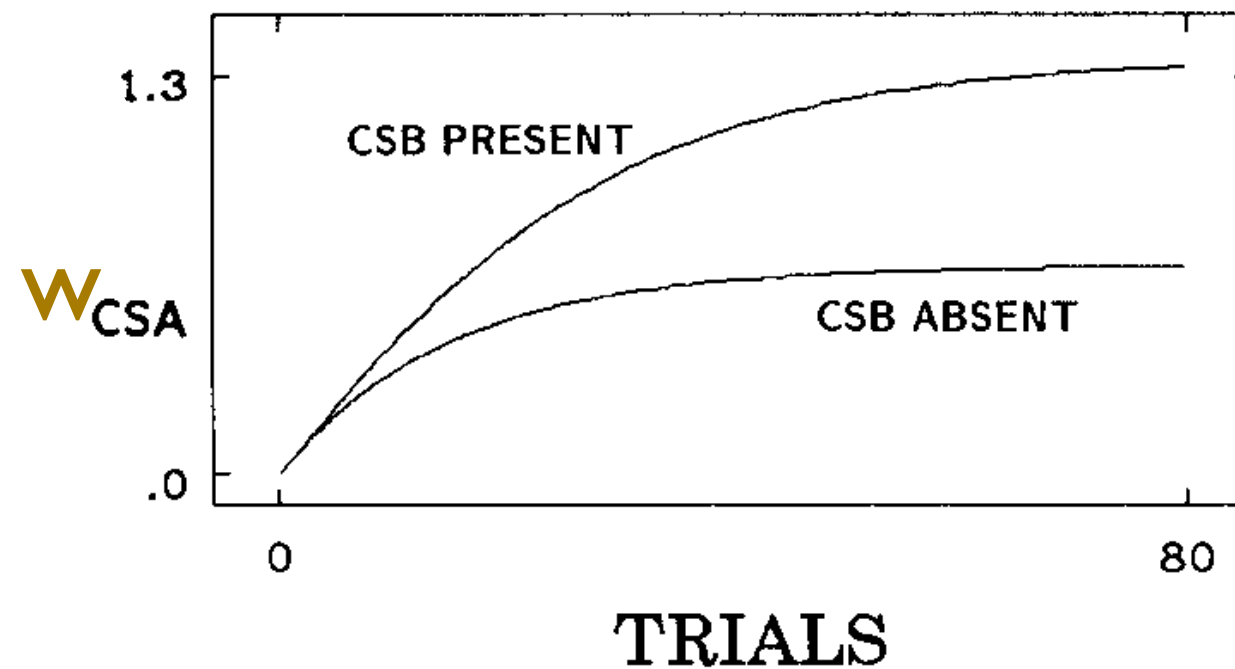
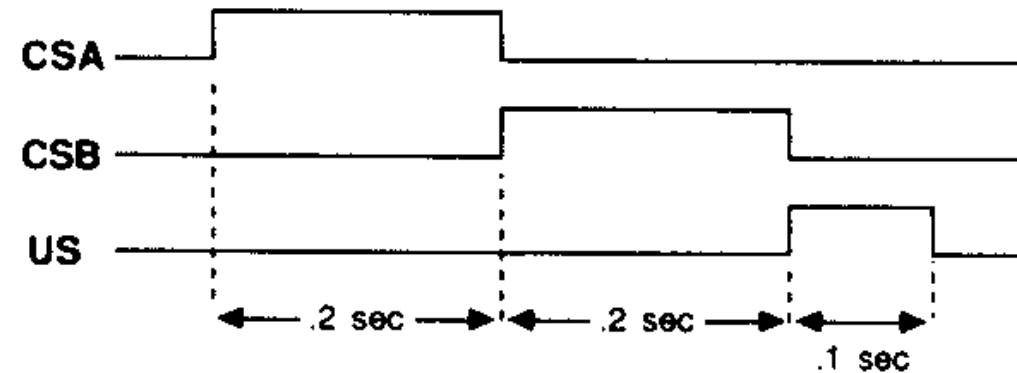
DELAY CONDITIONING



Second-order conditioning in the TD model

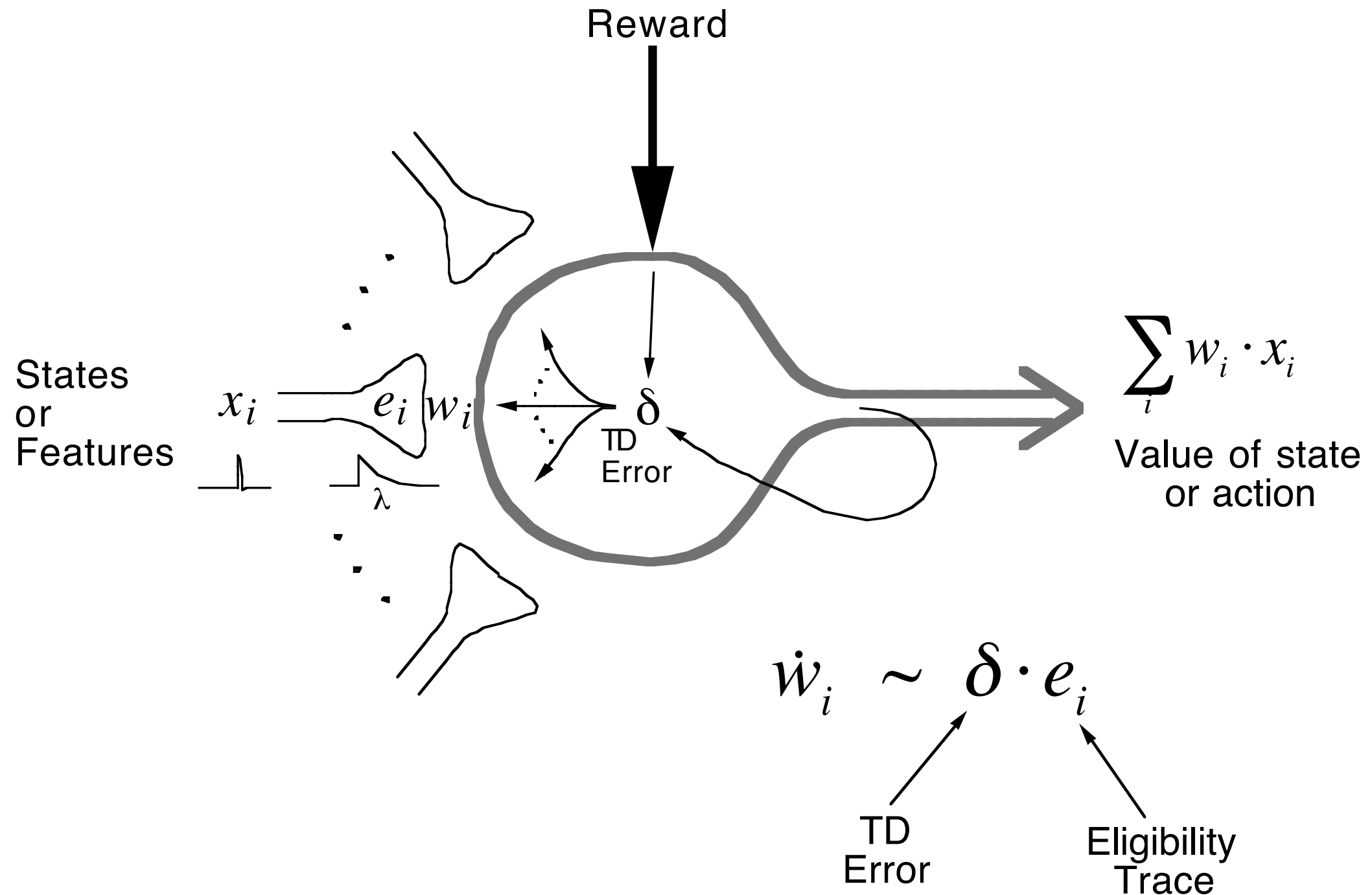


Primacy effect in the TD model

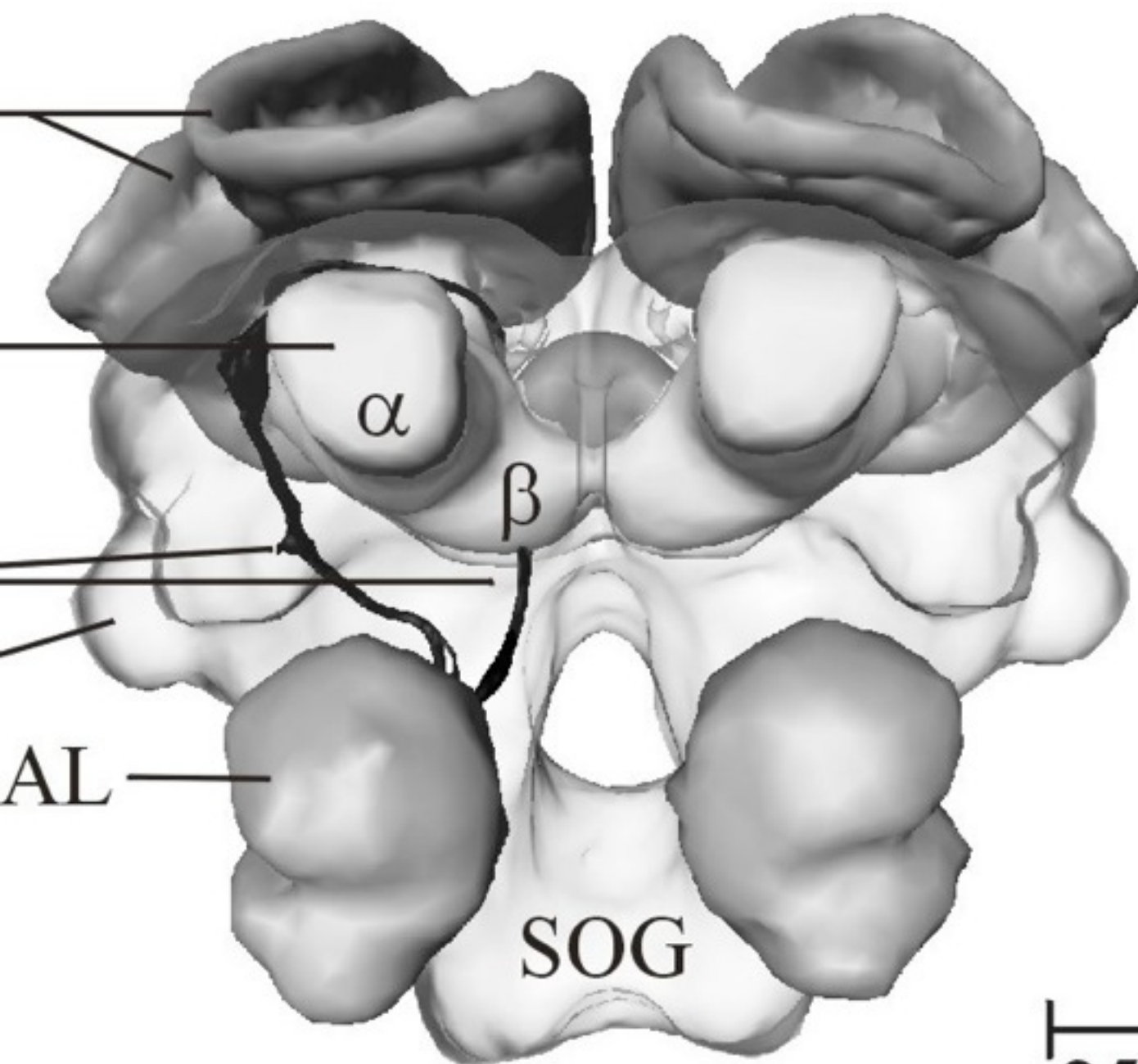


Facilitation of a remote association by an intervening stimulus

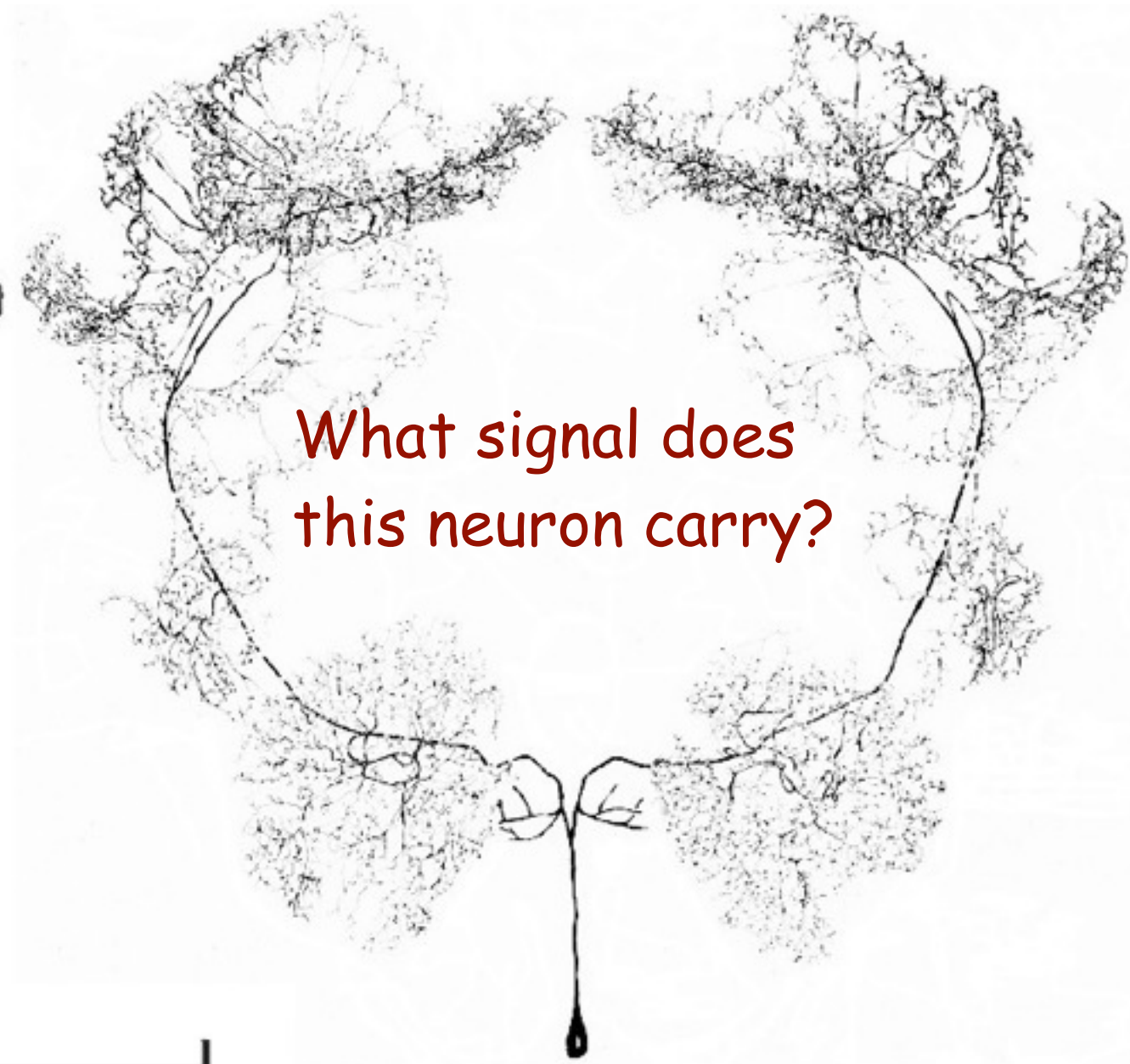
The TD model of classical conditioning as a single neuron



Brain reward systems



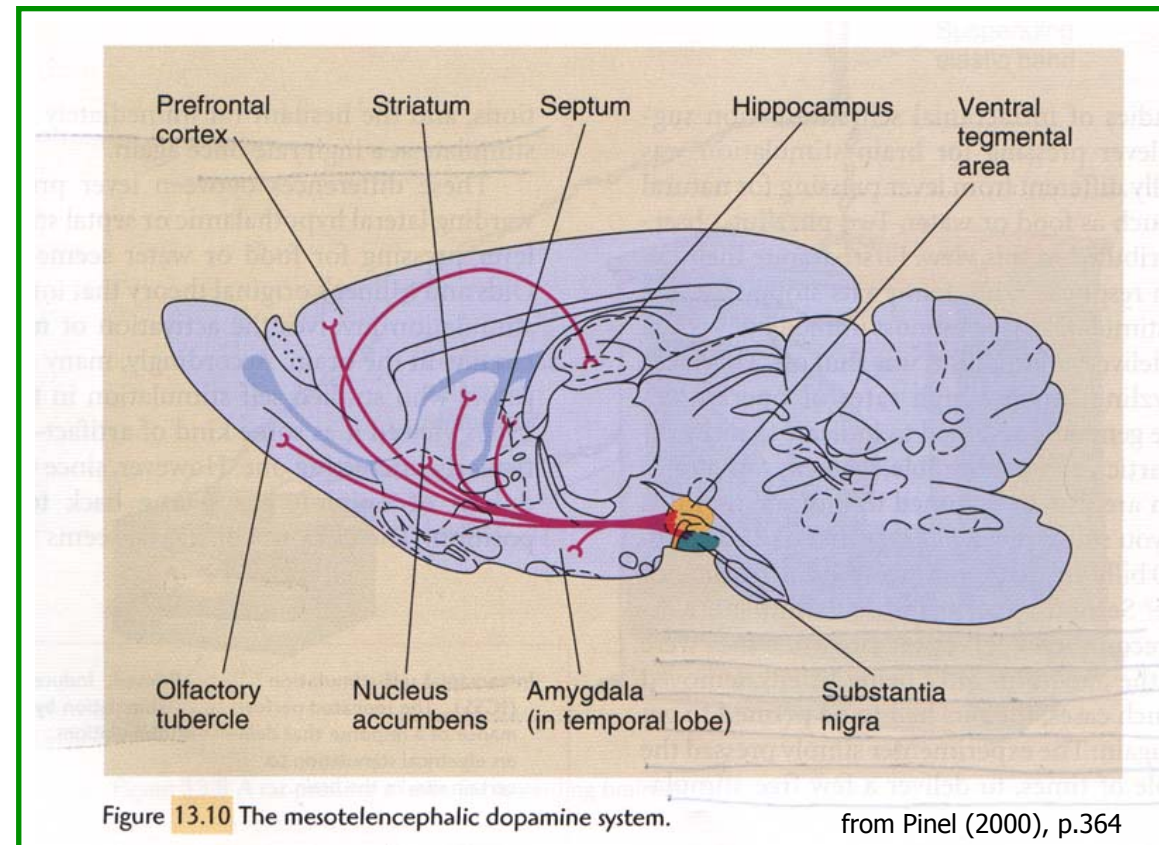
Honeybee Brain



VUM Octopamine
Neuron

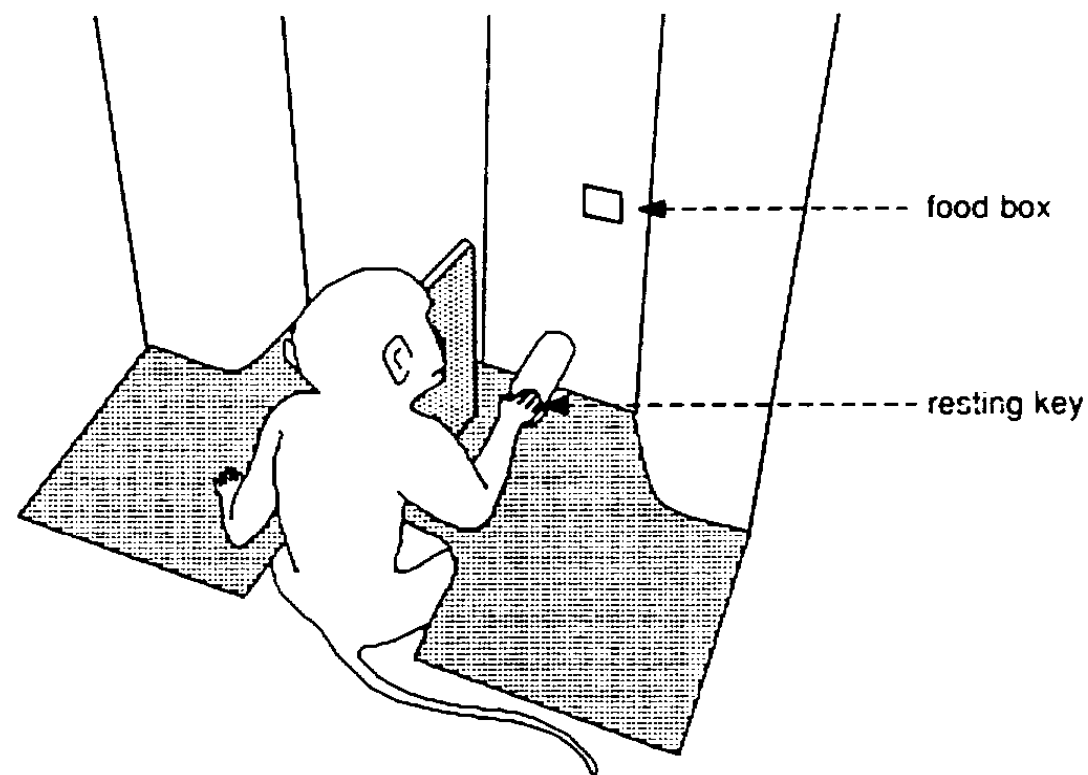
Dopamine

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



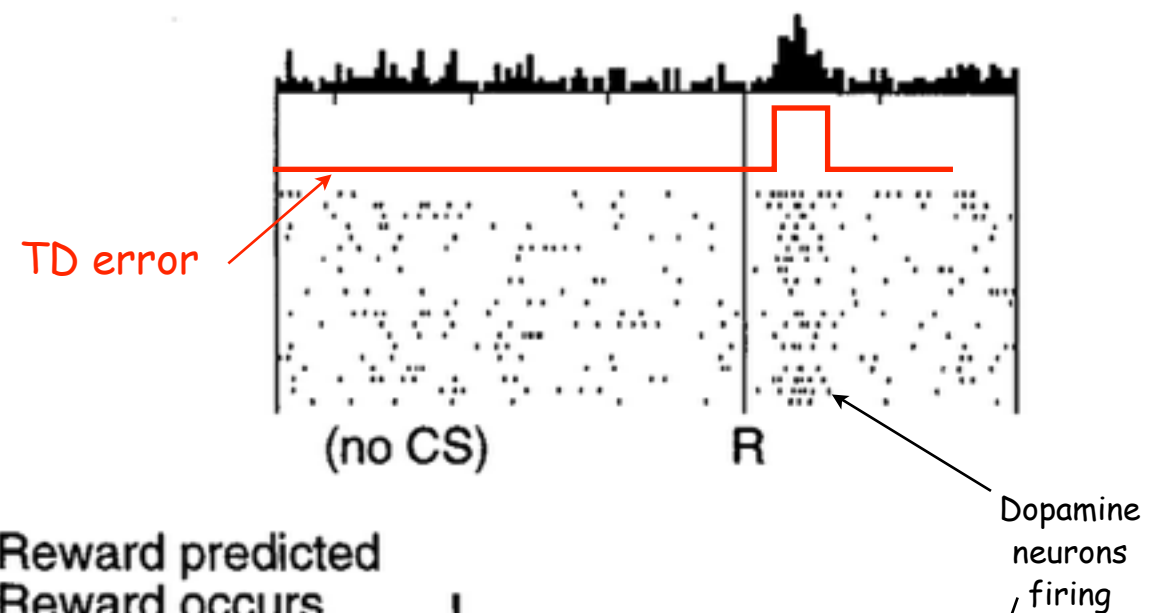
Key Idea: Phasic change in baseline dopamine responding = reward prediction error

Brain reward systems seem to signal TD error with dopamine

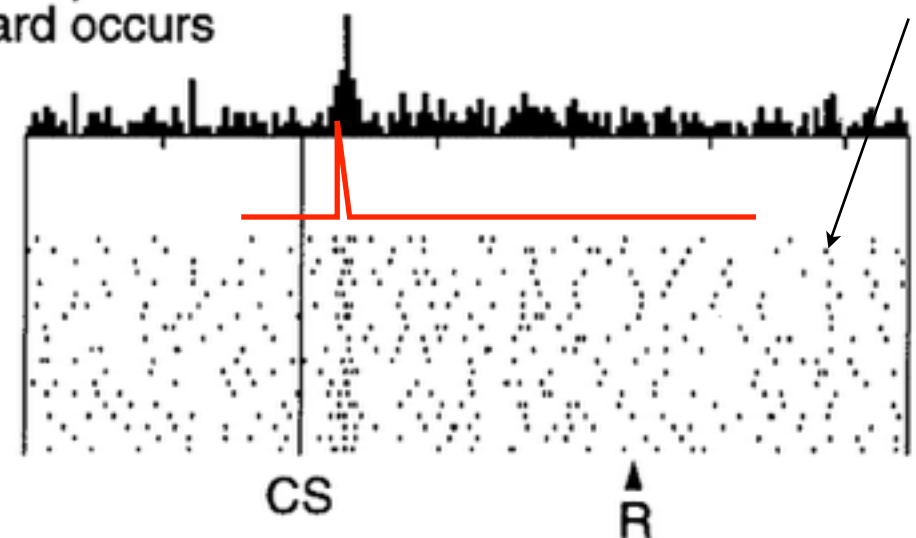


Wolfram Schultz, et al.

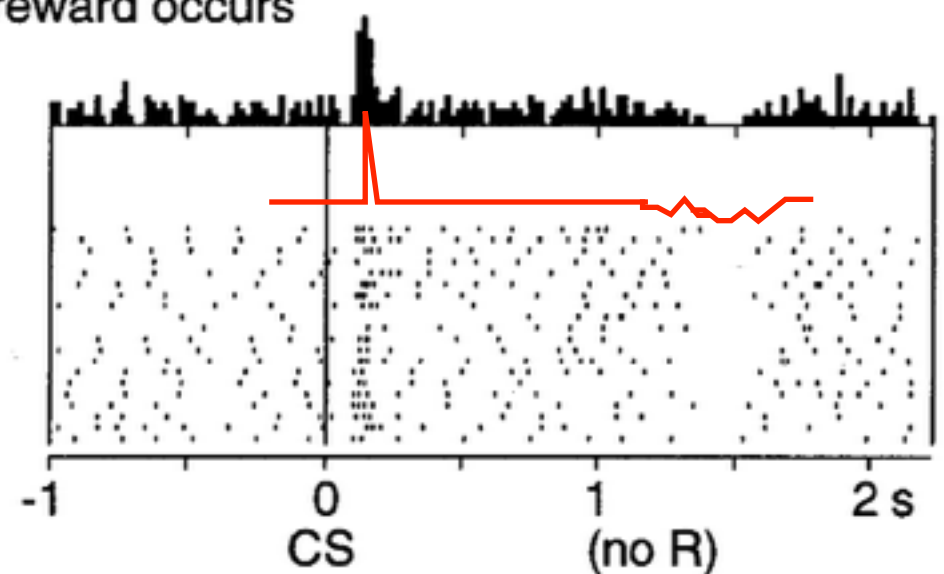
No prediction
Reward occurs



Reward predicted
Reward occurs



Reward predicted
No reward occurs



Theoretical TD Errors

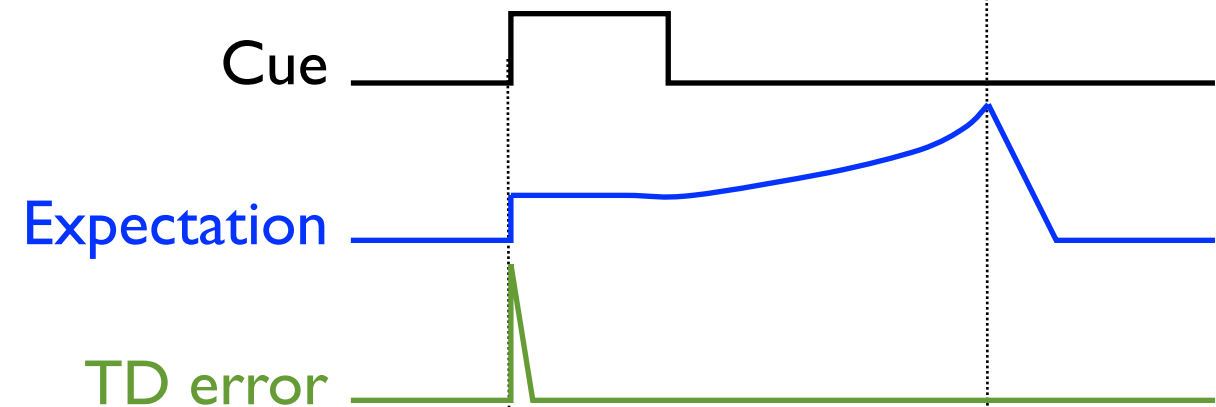
TD error:

$$\delta_{t+1} = R_{t+1} + .9E_{t+1} - E_t$$

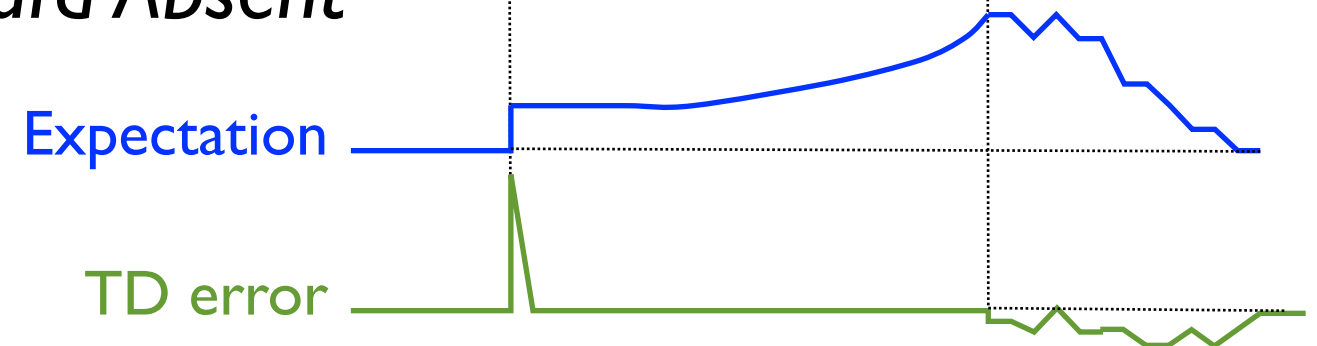
Reward Unexpected



Reward Expected



Reward Absent



Outline

1. The “discovery” of reinforcement learning

- that instrumental learning was missing from the engineering sciences

2. The discovery of temporal-difference learning

- in classical conditioning, as engineering, and in brain reward systems

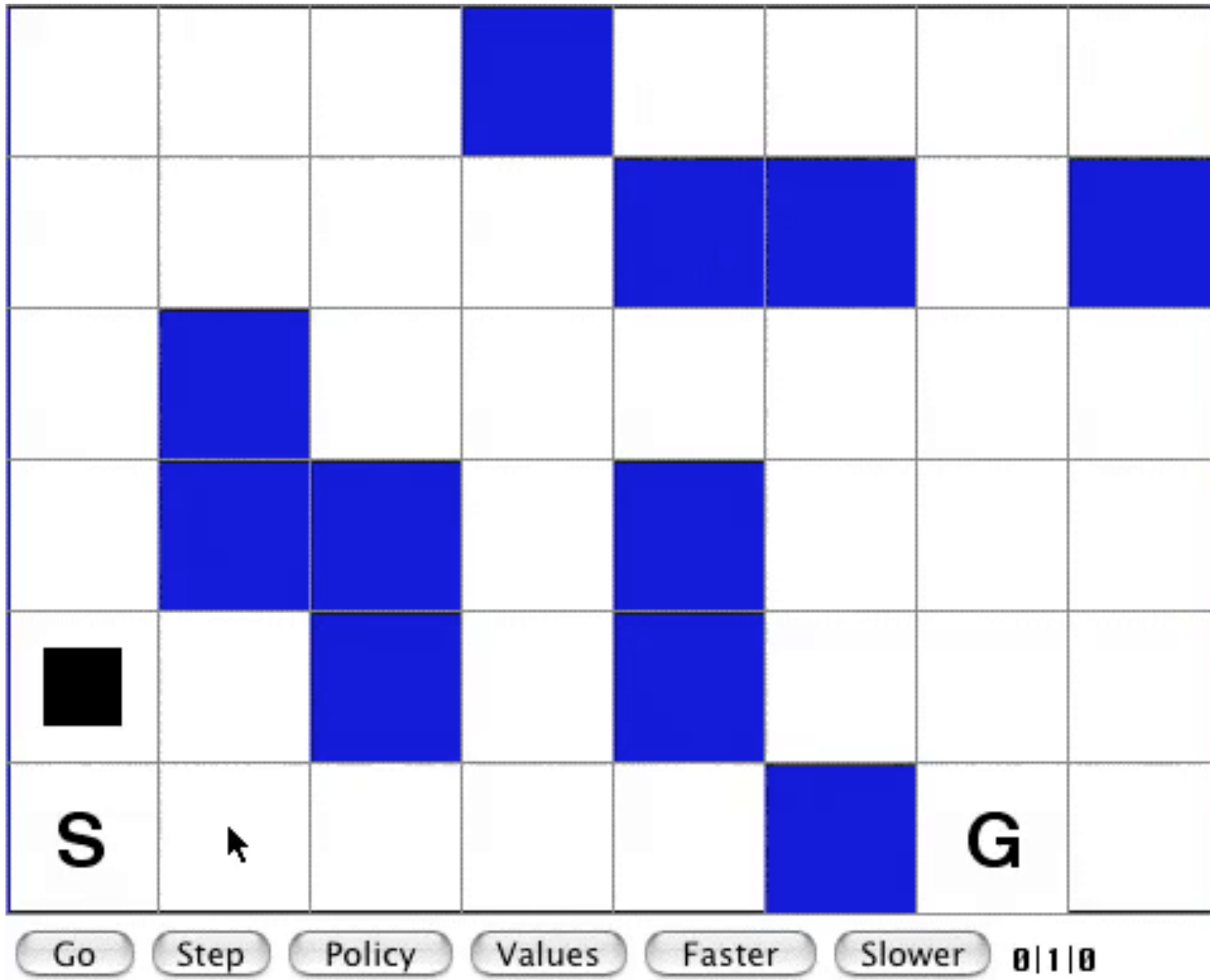
3. (Planning as RL on imagined experience)

Beyond trial and error

- There is an obvious simple strategy to achieve more cognitive abilities with RL methods:
 1. Learn a predictive model of the world
 2. Use it to generate imaginary experience
 3. Process the imaginary experience by RL algorithms *as if it were real*
- This has been explored in psychology, AI, and neuroscience

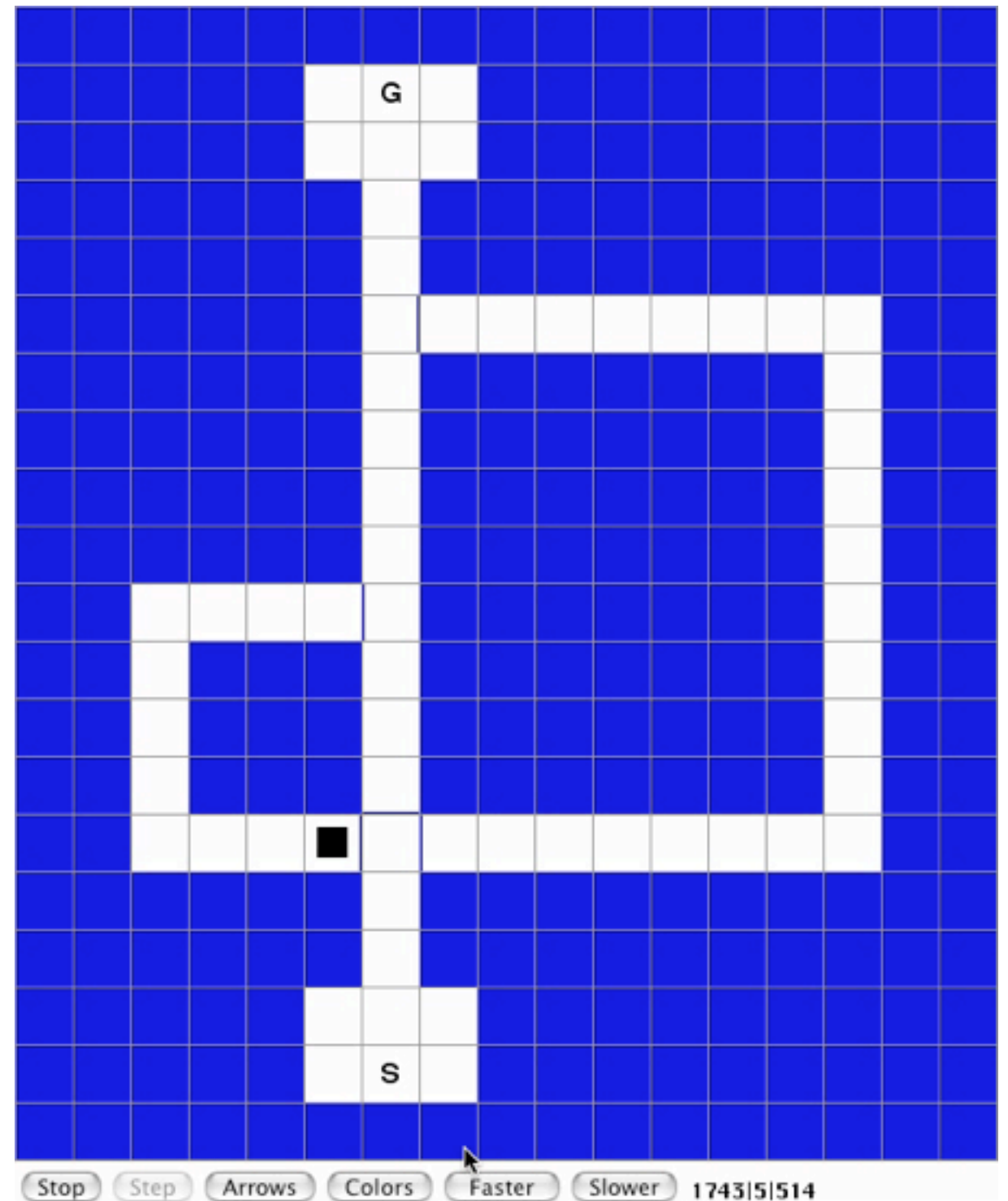
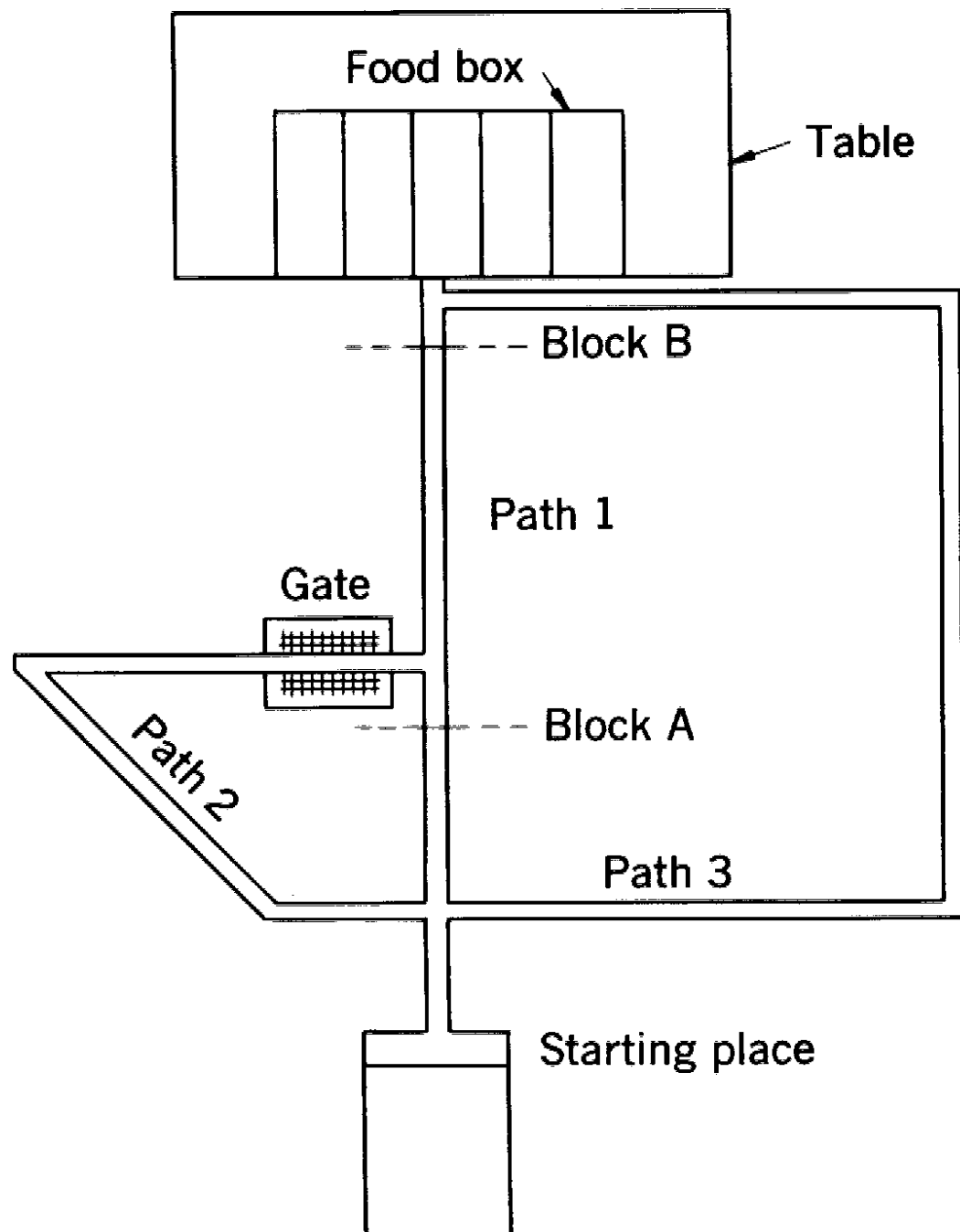
a form of planning, imagination, or even thinking

GridWorld Example



Tolman & Honzik, 1930

“Insight in Rats”



Marr's Three Levels at which any information processing system can be understood

- **Computational Theory Level**

- What are the goals of the computation?
- What is being computed?
- Why are these the right things to compute?
- What overall strategy is followed?

What and Why?

- **Representation and Algorithm Level**

- How are these things computed?
- What representation and algorithms are used?

How?

- **Hardware Implementation Level**

- How is this implemented physically?

Really how?

RL's computational theory of mind

- The overall goal is to maximize **reward**
- To max reward, we seek an optimal **policy**
- To help optimize the policy, we learn **expectations** of reward (a value function), and **TD errors**
- To do all of this with less data, we learn a **predictive model of the environment**, and apply the same methods to imagined experience

In conclusion:

What every psychologist should know about RL

1. RL is the engineering counterpart of instrumental learning (operant conditioning) in biology
2. RL propagates reward-prediction errors backward from goals (by TD methods)
3. Planning can be achieved by RL applied to replayed and imaginary experience
4. Psychology, AI, control theory and neuroscience are consilient; they may all flow together toward the same simple algorithms

Thank you for your attention

The RL&AI group at
the Univ. of Alberta
in 2011



Join us at the 2nd Multidisciplinary Conference on
Reinforcement Learning and Decision Making (RLDM)
on June 7-10, 2015, in Edmonton, Alberta, Canada