How simple can mind be?

Rich Sutton
Department of Computing Science
University of Alberta

Reinforcement Learning and Artificial Intelligence

rlai.net

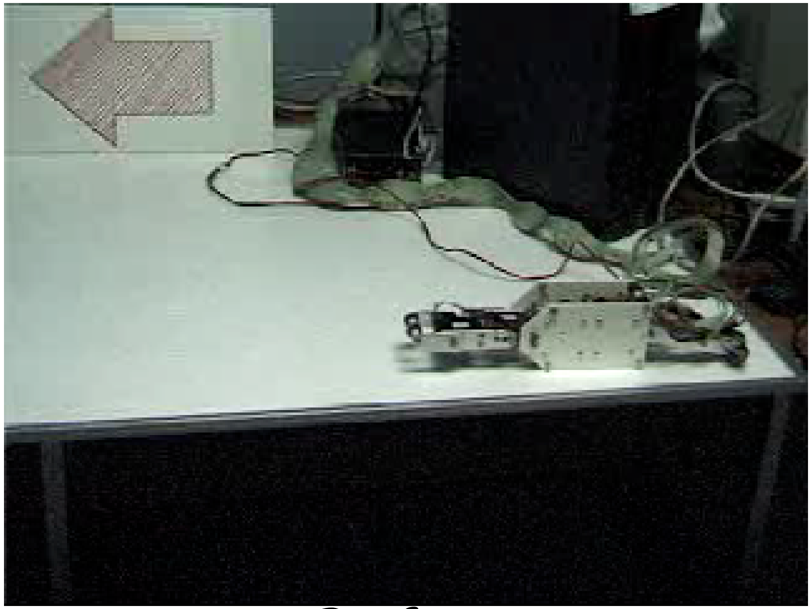
PIs:
Rich Sutton
Michael Bowling
Dale Schuurmans
Csaba Szepesvari

University of Alberta
Alberta Ingenuity Centre for Machine Learning
Informatics iCORE Circle of Research Excellence

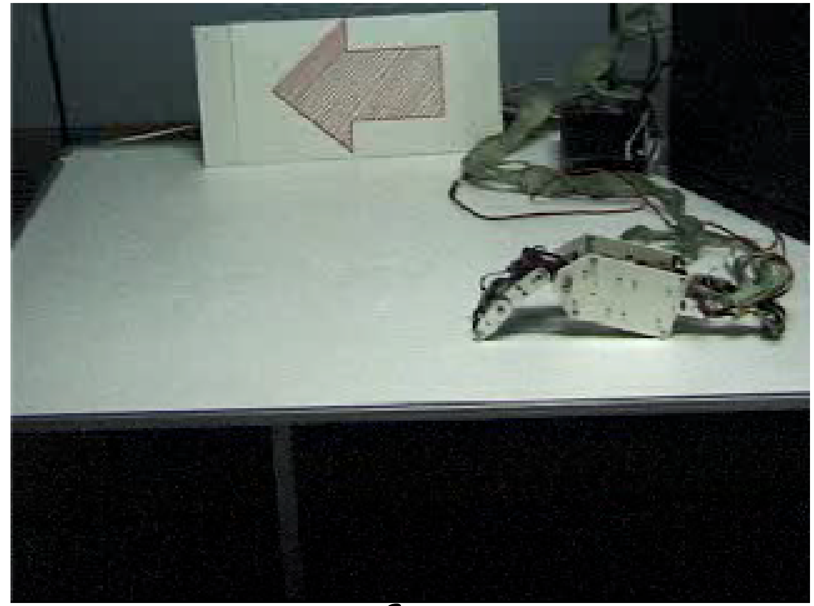
personal motivations

- To understand what mind means well enough to make some
- The incredible complexity of everyday knowledge and decision making
- Impatiently seeking general principles
  - reductionist, absolutist, simplistic
  - but ready to backtrack
- That horrible trial-and-error learning: Reinforcement learning

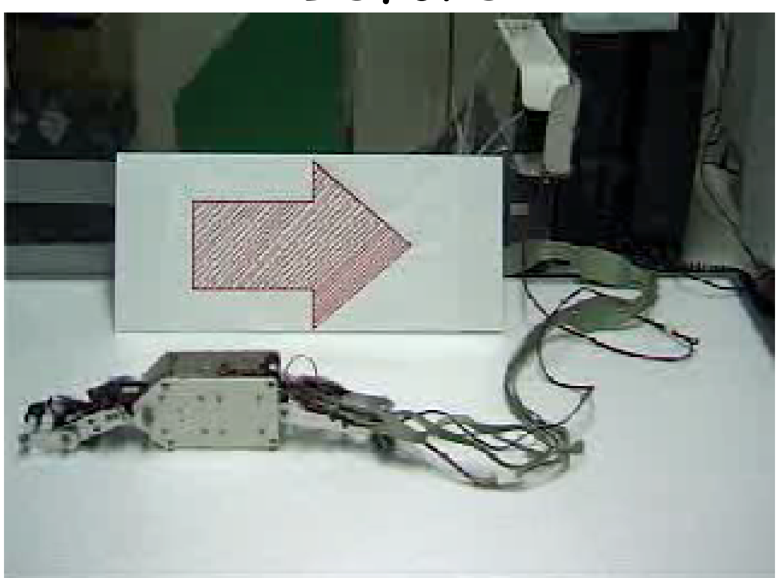
Hajime Kimura's RL Robots



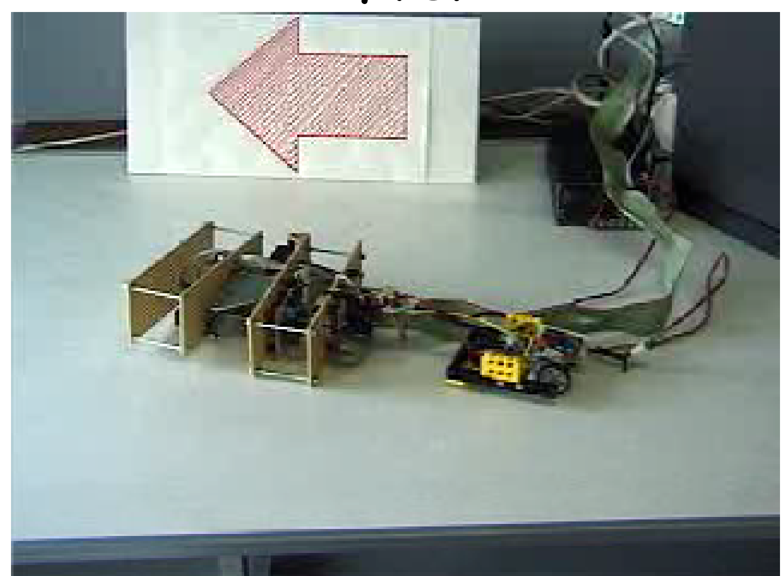
Before



After



Backward



New Robot, Same algorithm

AI is not biology

- AI is easier in some ways
  - we are more concerned with sufficiency
  - we know the agent's goal
  - we can look inside its head
  - we can ignore evolution
  - our experiments take less time
- On the other hand, we can't just theorize about mind – we have to actually make it

Marr's three levels of explanation for information-processing systems

- Computation theory
  - *What* is computed? — expected future reward
- Algorithms and representations
  - *How* is it computed? — TD learning
- Implementation
  - *Really*, how is it done? — TD error = Dopamine

Levels can be separated, validated independently

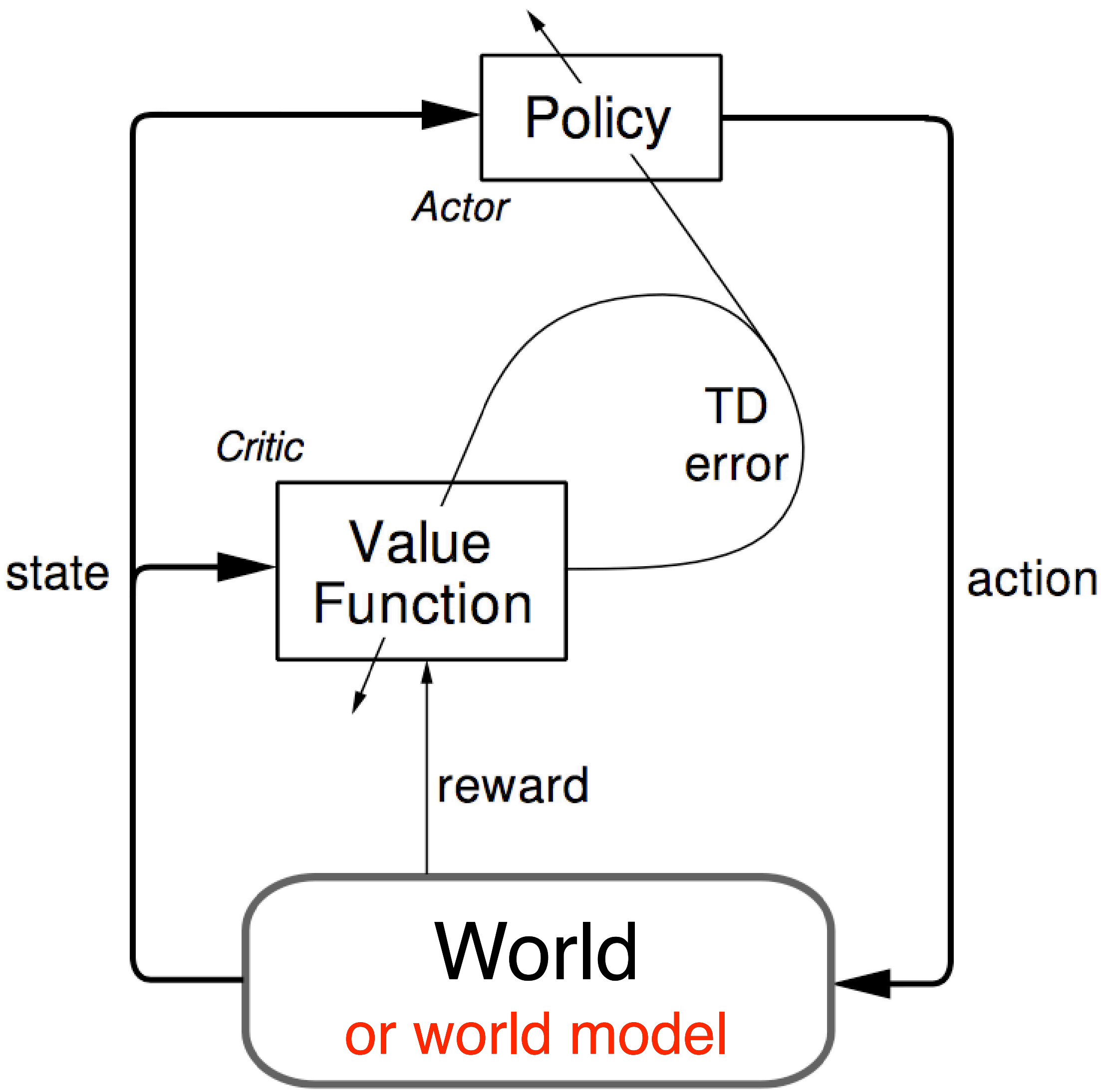
Ideas on offer

1. The interplay of goal-related signals: reward, value, and TD error
2. Learning on simulated experience (as planning, understanding, cognition, reasoning, thought, goal-directed...?)
3. Option models as an approach to the hard problem of representing knowledge that is abstract yet strongly linked to low-level experience

Essentials of mind (outline)

- Experience
- Goals
- Learning from experience
- Learning from simulated experience
- Abstraction
- Constructivism, discovery, generalization

Actor-critic architecture



S → R learning

S → S* learning

S → S learning

Understanding

- Knowing how the world works (having a predictive model of causes and effects)
- Being able to use that knowledge flexibly to achieve goals
  - a.k.a. planning, reasoning

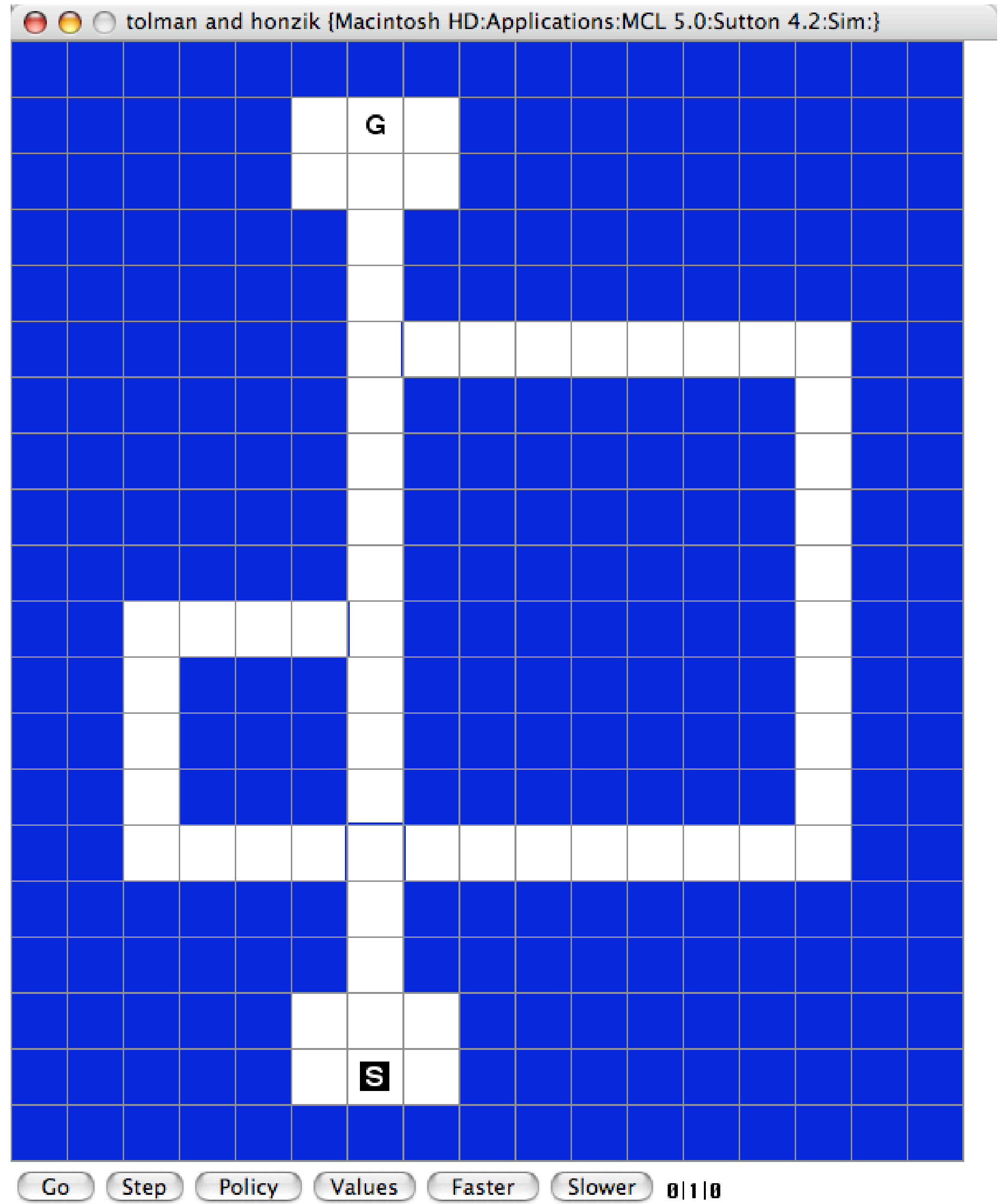
Learning from simulated experience

1. Learn a predictive model of the world
2. Use the model to generate simulated experience
3. Learn from the simulated experience as if it had actually happened

= cognition, model-based reasoning

cf. vicarious trial and error (Tolman, 1932)

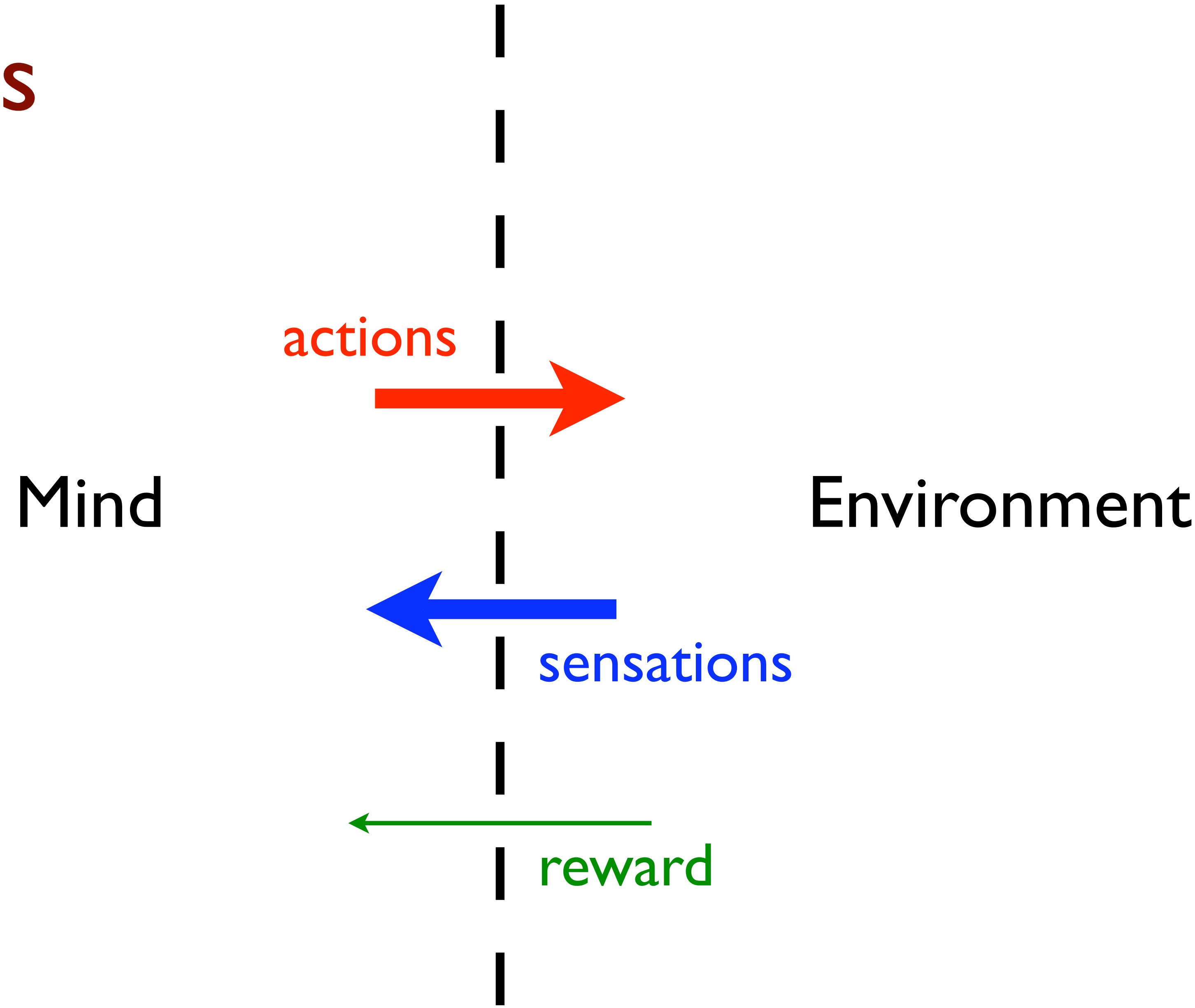
Recreation of Tolman & Honzik's "Reasoning in Rats" experiment (1930)



Experience

- The low-level stream of inputs and outputs – sensations and actions at 100Hz
- The final common paths of mind and world
- The data of artificial intelligence
- The only thing that is real
- It suffices to draw a hard line...

Goals



The reward hypothesis

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

- Simple, but not trivial
- A good *null hypothesis*

Values

- A value V_t is an expectation of cumulative future reward:

$$V_t = E\left\{\sum_{k=1}^{\infty} r_{t+k}\right\}$$

- Values are defined in terms of rewards
- Approximate values $\hat{V}_t \approx V_t$ are learned from experience
- Rewards are primary, values secondary
  - but it is values that guide decision-making

The value hypothesis

All efficient methods for solving sequential decision problems must learn or compute values as an intermediate step

- dynamic programming
- most reinforcement learning methods

TD error

- For learning, the key scalar is neither reward nor value, but the temporal-difference error:

$$r_{t+1} + \hat{V}_{t+1} - \hat{V}_t$$

- The TD error is a measure of how pleased or disappointed you are in moving from t to $t+1$:

$$\begin{aligned}
\hat{V}_t &\approx \sum_{k=1}^{\infty} r_{t+k} \\
&= r_{t+1} + \sum_{k=2}^{\infty} r_{t+k} \\
&\approx r_{t+1} + \hat{V}_{t+1}
\end{aligned}$$

- The interplay of reward, value, and TD error is a significant contribution to our understanding of goal-directed learning

Abstraction in time and state

- Options
  - a way of behaving with a termination condition
- Option models
  - predictions about the outcomes of options
- Compositionality
  - predictions can be about other predictions

Examples of option-conditional compositional predictions

If I were to follow this hallway to its end,
would I find a restroom?

If I were to look in the fridge,
would I see a beer?

If I were to open the box,
would I see an apple?

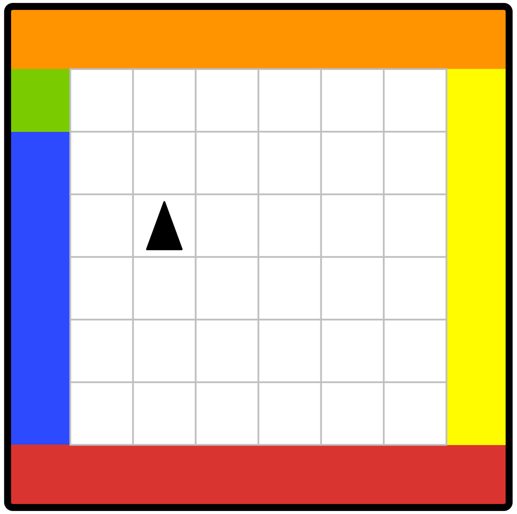
If I were to turn over the glass,
would the carpet be wet?

Outcomes are not primitive observations

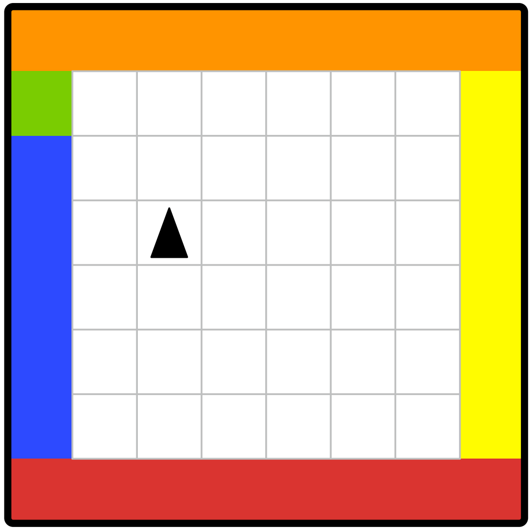
They are sets of predictions

Compass world

- sensation: color ahead
- actions:
  - L(eft)
  - R(ight)
  - F(orward)
- options:
  - Leap (to wall)
  - Wander (randomly)



Examples in compass world



If I were to...

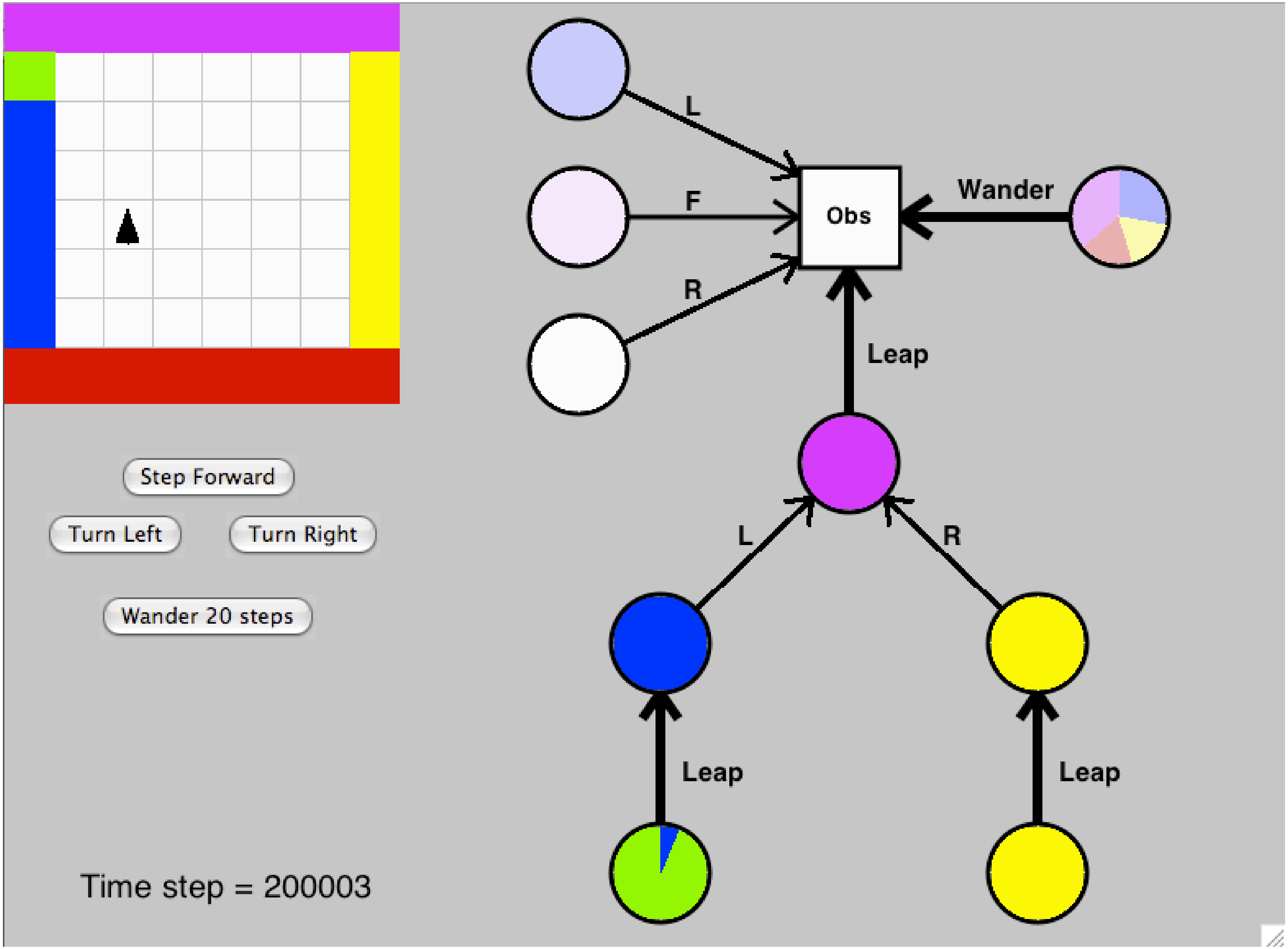
...step forward till I hit a wall,
would it be orange?

"facing an orange wall"

not compositional

...step forward till I hit a wall, then turn left,
would I be "facing a green wall?"

compositional


L
F
R
Obs
Wander
Leap
Step Forward
Turn Left
Turn Right
Wander 20 steps
L
R
Leap
Leap
Time step = 200003

Constructivism a.k.a. discovery

- We have machinery for representing abstract knowledge
- We have so-so algorithms for learning option models
- But we don't know how to automatically create options with good properties:
  - Markov, linear, independent, not too numerous
- We construct the world
  - and I have no idea how

Conclusions

- Simple general principles are possible in AI
  - that may relate to animal behavior
- Learning from simulated experience suffices to explain much that seems beyond ordinary associative learning
- RL's sense of reward, value, and TD error contribute to understanding goal-directed behavior
- It may be possible someday to relate abstract knowledge to low-level experience