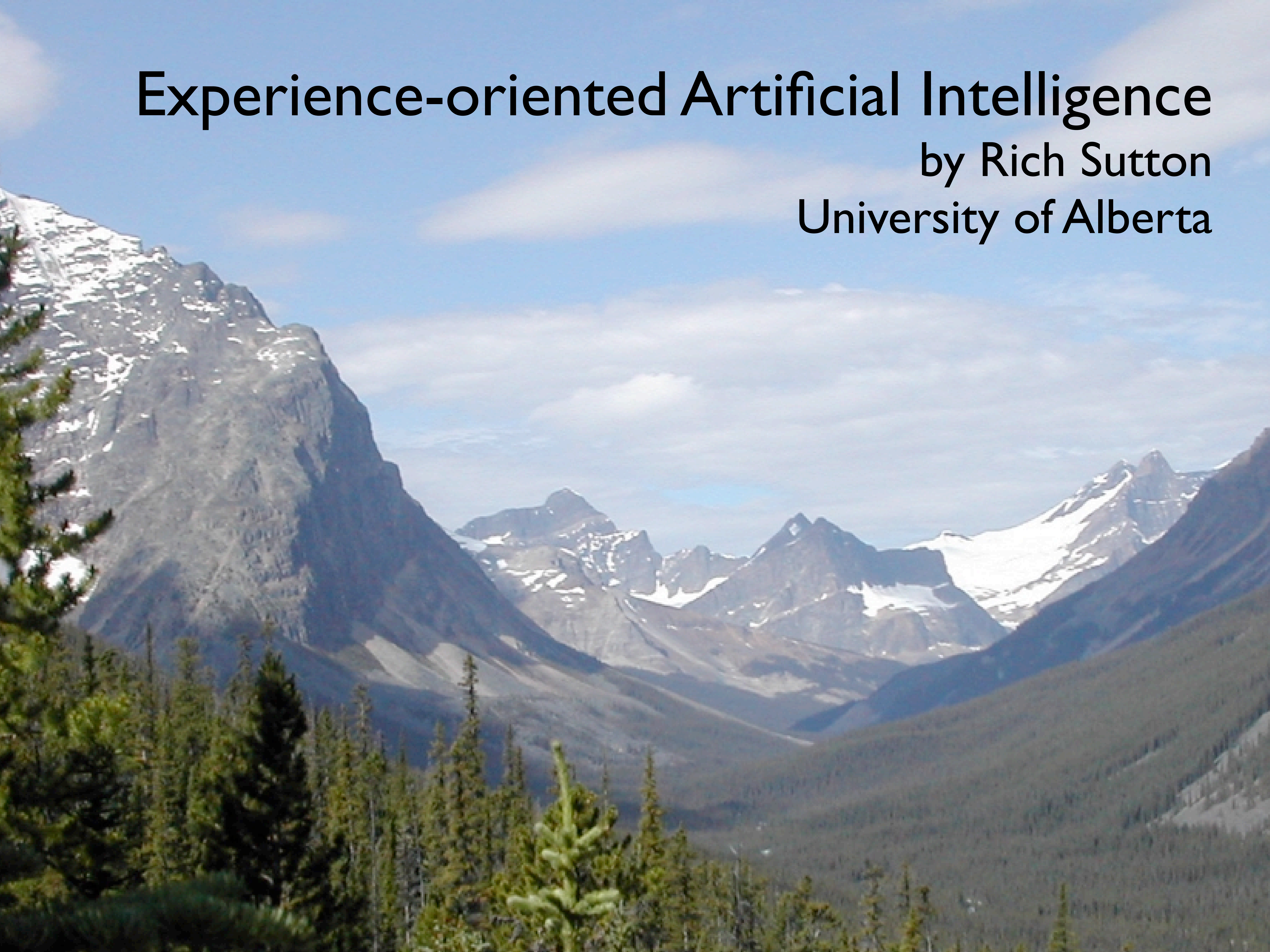
Experience-oriented Artificial Intelligence

by Rich Sutton
University of Alberta



Reinforcement Learning and Artificial Intelligence

PIs:
Rich Sutton
Michael Bowling
Dale Schuurmans

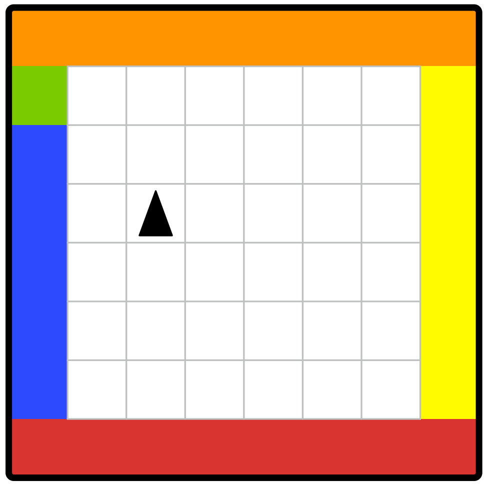
University of Alberta
Quaecumque Vera

Alberta Ingenuity Centre for Machine Learning

Informatics iCORE Circle of Research Excellence

Outline

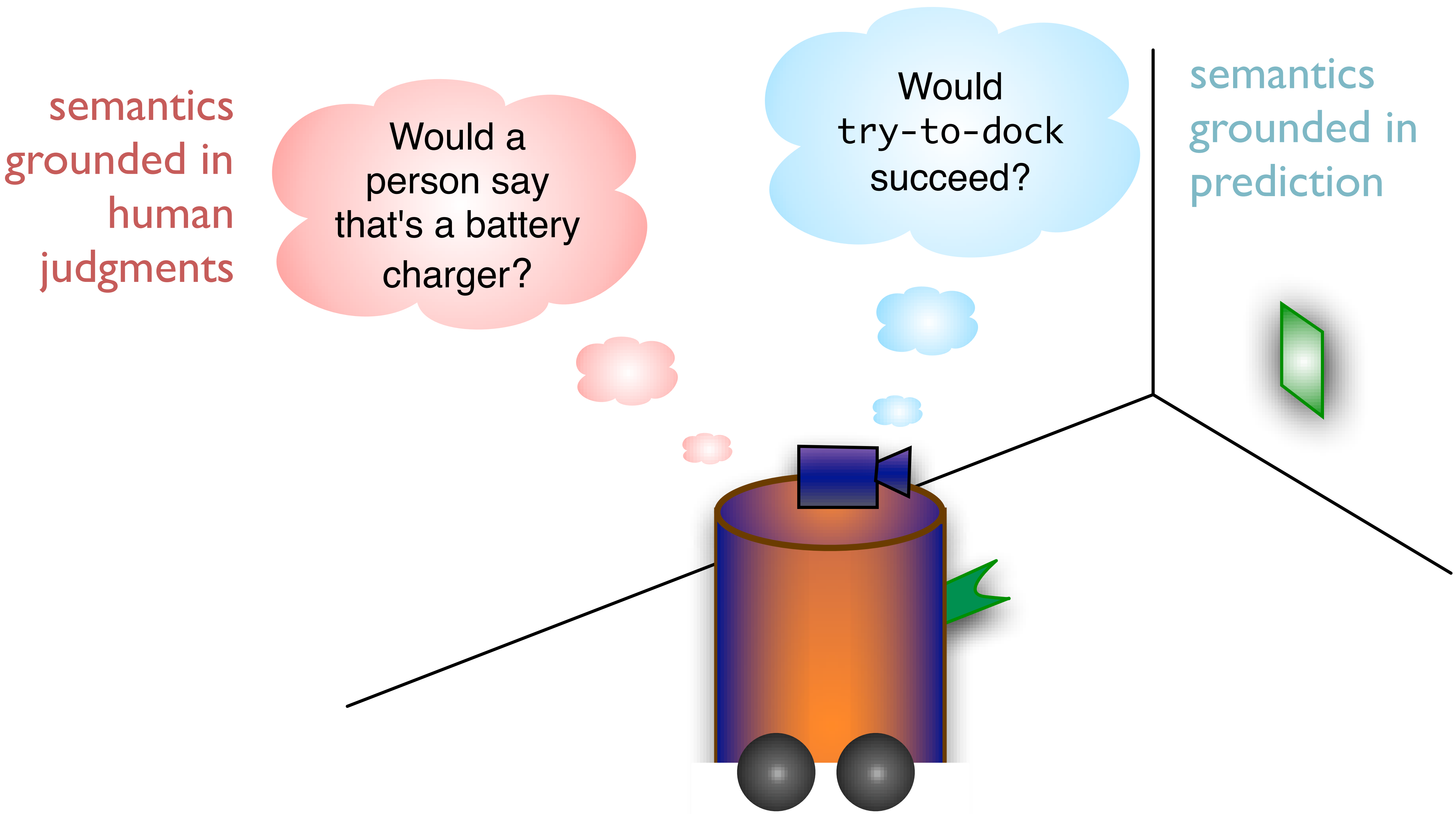
- Perspective on AI
- A predictive conception of world knowledge
- Machinery for predictive knowledge
  - options, PSRs, TD networks
- Micro-world experiments

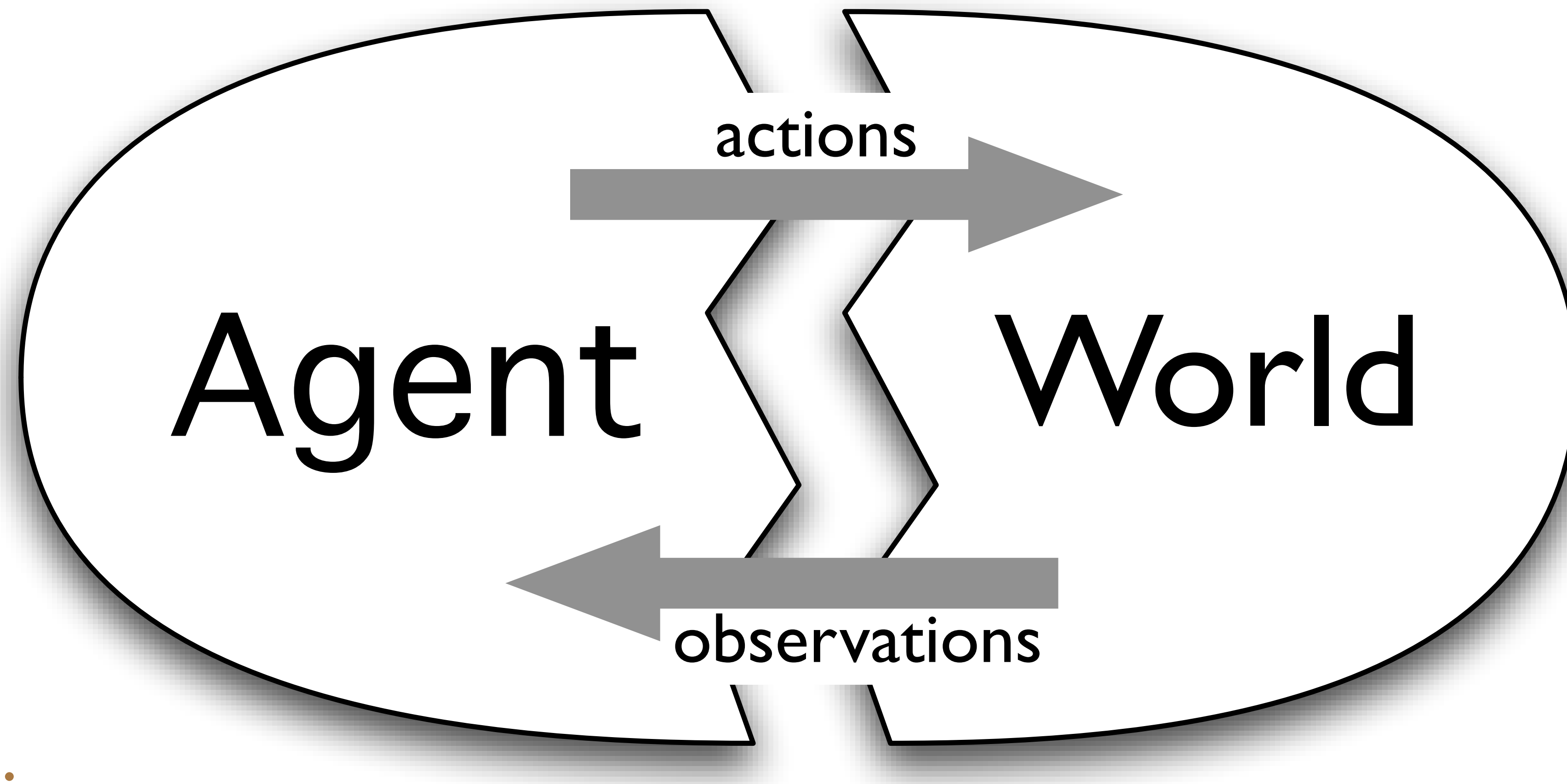


Take-home messages

- AI should be oriented around experience
  - but it's not
- Knowledge must be predictions
  - but that's nearly unimaginable
- Predictions can be really complex, abstract, expressive and compositional
  - while their machinery is simple and uniform
- Run-time verification may enable big AI
  - although I will show you just small AI

Run-time verification is the key to AI





observation

action

$$o_1\, a_1\, o_2\, a_2\, o_3\, a_3\, o_4\, a_4\, o_5\, a_5\, o_6\, a_6\, o_7\, a_7 \cdots$$

past

now

future

low-level sensori-motor experience, e.g., 100 Hz

Experience matters

- Experience is the most prominant feature of the computational problem we call AI
- It's the central data structure
- It has a definite temporal structure
  - revealed and chosen over time
  - speed of decision is important
  - order is important
- This has unavoidable implications for AI

Experience in AI

Many, many AI systems have no experience

They don't have a life!

- Expert Systems
- Knowledge bases like CYC
- Question-answering systems
- Puzzle solvers,
  - or any planner that is designed to receive problem descriptions and emit solutions

Part of the new popularity of agent-oriented AI is that it highlights experience

Other AI systems have experience, but don’t focus on it

Orienting around experience suggests radical changes in AI

Knowledge of the world should be
knowledge of possible experiences

Planning should be about
foreseeing and controlling experience

The state of the world should be
a summary of past experience,
relevant to future experience

Take-home messages

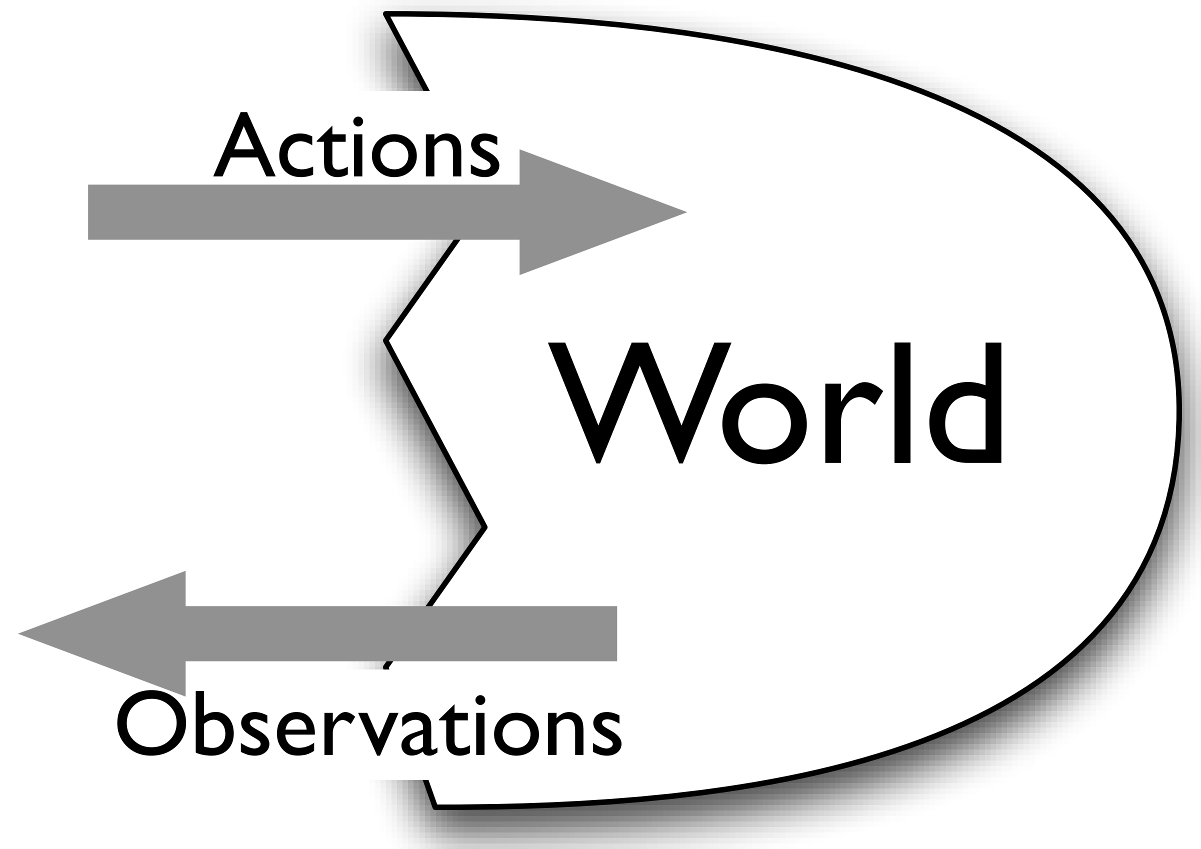
- AI should be oriented around experience
  - but it's not
- Knowledge must be predictions
  - but that's nearly unimaginable
- Predictions can be really complex, abstract, expressive and compositional
  - while their machinery is simple and uniform
- Run-time verification may enable big AI
  - although I will show you just small AI

World knowledge must be predictions

actions $a_t \in \mathbf{A}$

observations $o_t \in \mathbf{O}$

experience $e_t \in \{\mathbf{O} \times \mathbf{A}\}^t$



The world is completely described by the probability distribution

$\omega(o \mid e) = \mathrm{Prob}(o_{t+1} = o \mid e_t = e)$

To know something about the world at time t

is to know something about $\omega(o \mid e_t e)$ for $e \in \{\mathbf{O} \times \mathbf{A}\}^*$

There is nothing else to know

- Everything we know that is specific to this world (as opposed to universally true in any world) is a prediction or memory of experience
- All world knowledge must be translatable into statements about future experience

A Grand Challenge

- To represent human-level world knowledge solely in terms of
  - observations (includes rewards, if any)
  - actions
  - time steps
- without reference to any other concepts or entities unless they are themselves represented in terms of experience

What would it be like to accept the challenge?

- Abstraction is key
  - state
  - dynamics
- Need to think in unfamiliar ways
- Microworlds, robotics
- Indexical (deictic) representations
  - sequence instead of symbols

In experential terms,

- What is space?
  - regularities in sensation change with eye movement
- What are objects?
  - subsets of sensations
  - that tend to occur together temporally
  - and can be in arbitrary relative spatial arrangements

- What is my body, my hands?
  - objects that are always present
  - and can be controlled
- What are people?
  - objects that may move on their own
  - that have a particular subset of sensations
  - whose presence may change my sensations for the better
  - eventually:
    - ✦ that are best predicted with respect to goals
    - ✦ that are analogous to me

What would it be like to accept the challenge?

- Abstraction is key
  - state
  - dynamics
- Need to think in unfamiliar ways
- Microworlds, robotics
- Indexical (deictic) representations
  - sequence instead of symbols

Philosophical and Psychological Roots

- Like classical british empiricism (1650–1800)
- Like logical positivism (Ayer, Peirce)
- But not anti-nativist, not tabula rasa
- Subjective rather than objective
- Emphasizing sequential rather than simultaneous events
- Close to Tolman's "Expectancy Theory" (1932–1950)
  - Cognitive maps, vicarious trial and error

Take-home messages

- AI should be oriented around experience
  - but it's not
- Knowledge must be predictions
  - but that's nearly unimaginable
- Predictions can be really complex, abstract, expressive and compositional
  - while their machinery can be simple and uniform
- Run-time verification may enable big AI
  - although I will show you just small AI

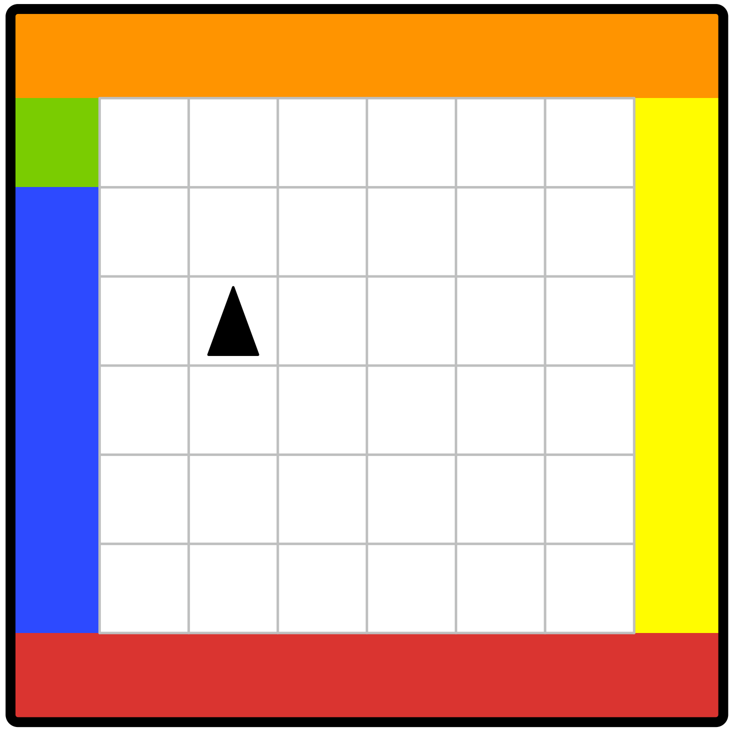
Sutton, Precup & Singh, 1999

Key machinery 1: options

- options are a generalization of actions
  - a way of behaving (policy), $\pi : \mathbf{S} \times \mathbf{A} \rightarrow [0,1]$
  - a way of stopping (term. cond.), $\beta : \mathbf{S} \rightarrow [0,1]$
- for the robot and the battery charger:
  - behave according to some try-to-dock policy
  - stop when docked or timed out

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...

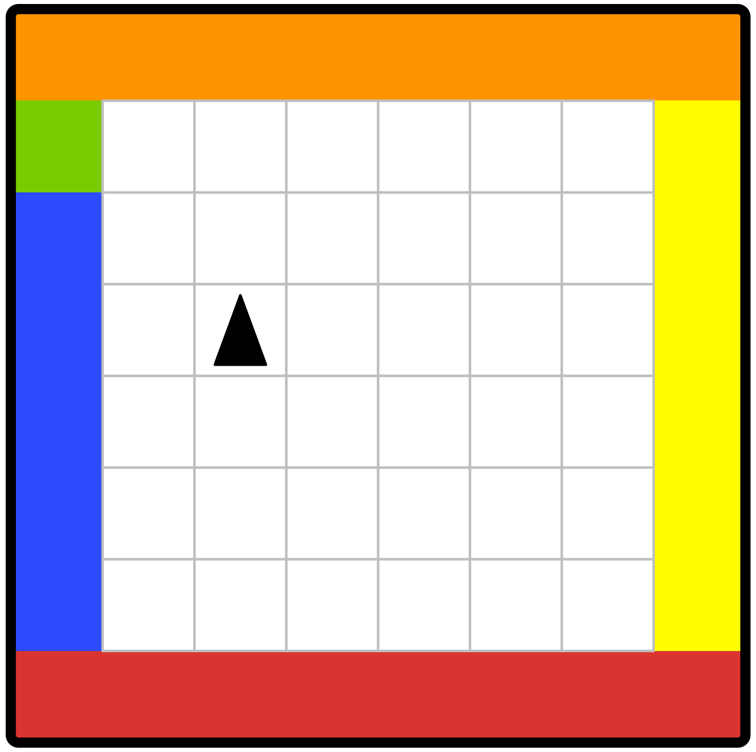
...step forward till I hit a wall,
would it be orange?

"facing an orange wall"

not compositional

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

compositional



Why options?

- they are very simple and general
  - a minimalist, least-commitment form of macro-action
  - allow arbitrary closed-loop policies
  - support action-independent temporal abstraction
- they are compatible with planning methods based on dynamic programming

Key machinery 2: option models

- an option model is a prediction of the option's outcome
  - what state you will end up in: $p : \mathbf{S} \times \mathbf{S} \rightarrow [0,1]$
  - how much reward you'll get along the way: $r : \mathbf{S} \rightarrow \mathbf{R}$
- for the robot and the battery charger:
  - will I end up docked?
  - will it hurt along the way, or take a long time?
- These are *subjunctive* predictions – "If I were to..."

Examples of subjunctive, compositional predictions

If I were to...

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

...look in the fridge,
would I see a beer?

...open the box,
would I see an apple?

...turn over the glass,
would the carpet be wet?

Outcomes are not primitive observations

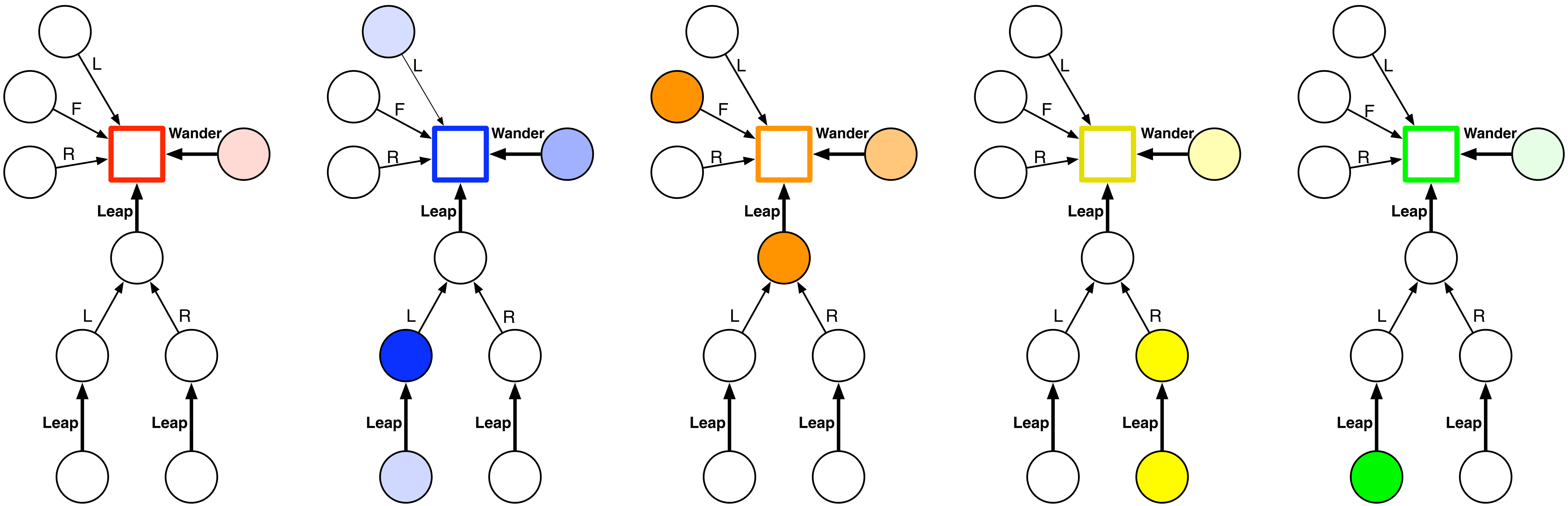
They are sets of predictions

Littman, Sutton, Singh, 2001

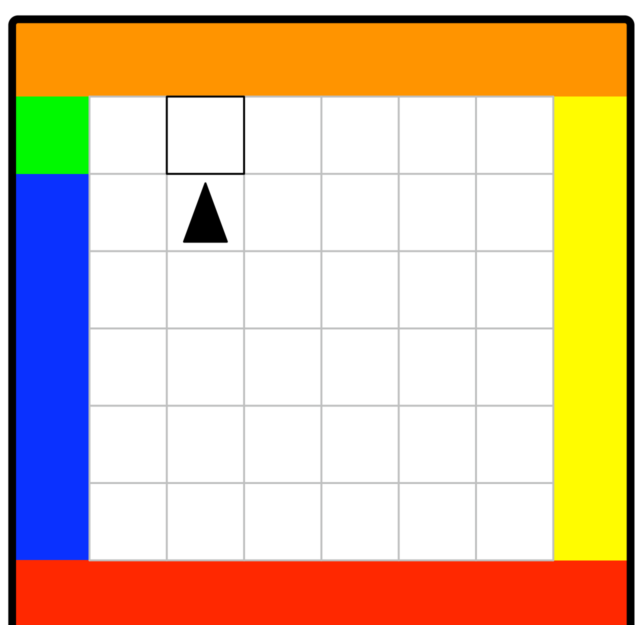
Key machinery 3: Predictive representations of state

- Use predictions of option outcomes as state variables
- for the robot and the battery charger:
  - is this a state where try-to-dock will succeed? a.k.a. is there a battery charger here?
  - is this a state where roll-backwards will trigger my bump sensor? a.k.a. is there an obstacle behind me?

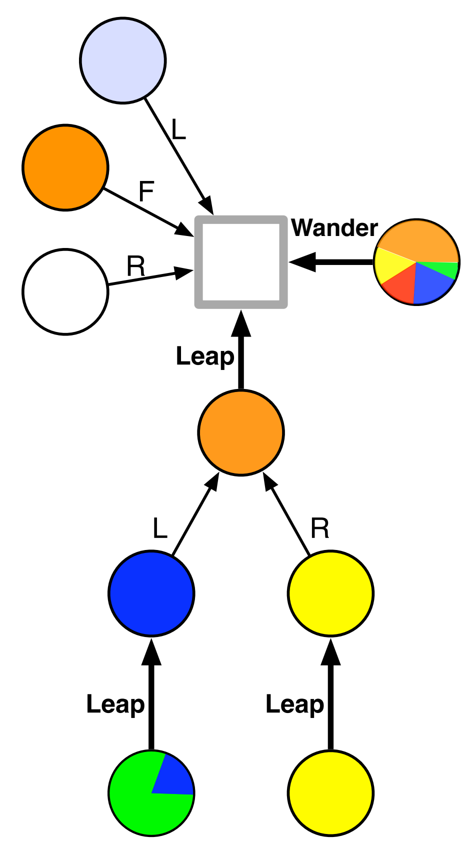
Complete question network



For this
world state



=



State is thus exorcised

- State is reduced to predictions of experience
- Option models are usually state to state
- Now they are state variable to state variable
- And the state variables are predictions
  - may be direct predictions of experience
  - or may be predictions of other predictions – *compositionality*

Temporal-difference networks

- Represent state and knowledge as predictions of predictions
- Divide the problem of prediction into two parts
  - specifying the questions about the future
  - computing their answers
- One set of nodes, two sets of interconnections

Answers are relatively easy to represent; it's questions that are hard

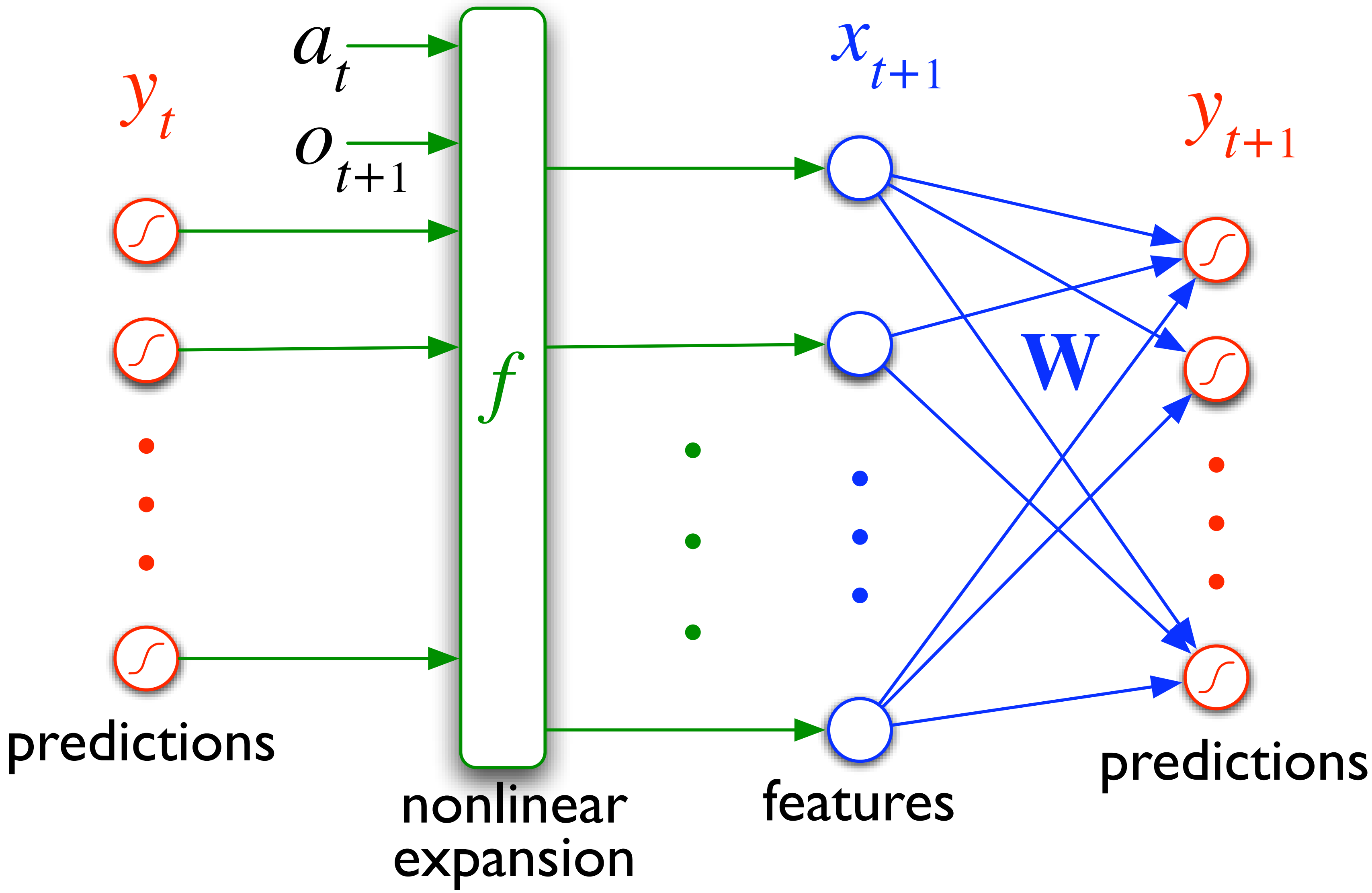
- e.g., flipping a coin
  - Question: what is the probability of heads
  - Answer: 0.5
- How to represent *flipping*, *coin*, and *heads*?

- What is *heads*?
- It's *not* a sensation
- It's another *prediction*
- We need to be able to ask questions about *predicting predictions*
- We need *compositionality*
  - predictions that can be built out of other predictions
- We need *abstraction*
  - predictions that capture similarities

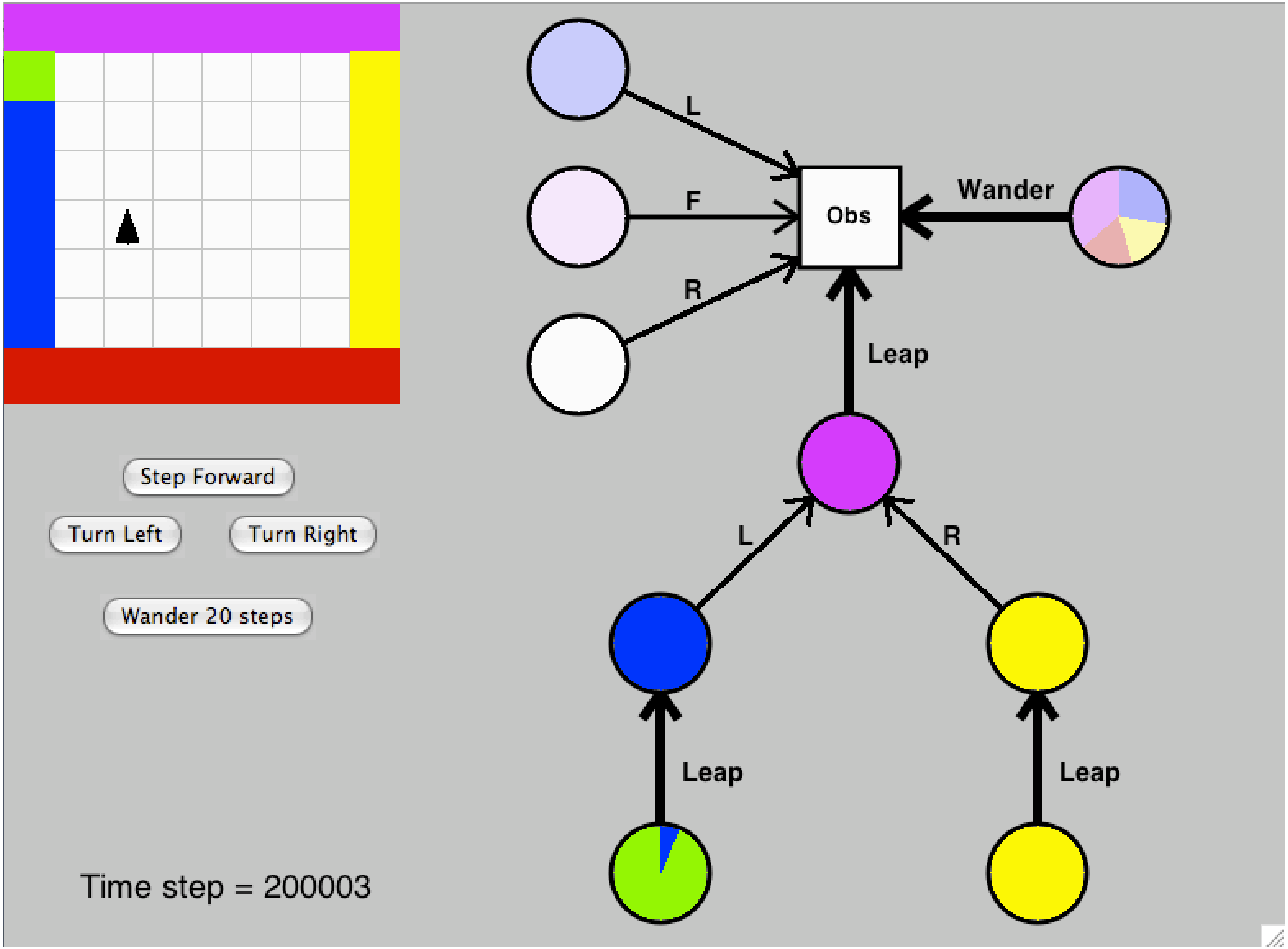
Qs & As in TD nets

- Answers are scalars
- Questions are "What would be the value of this signal at the end of this option?"
  - question = target signal, option
    = z, π, β
  - the target is often the answer to another question

Answer network structure



Answer networks compute the predictions

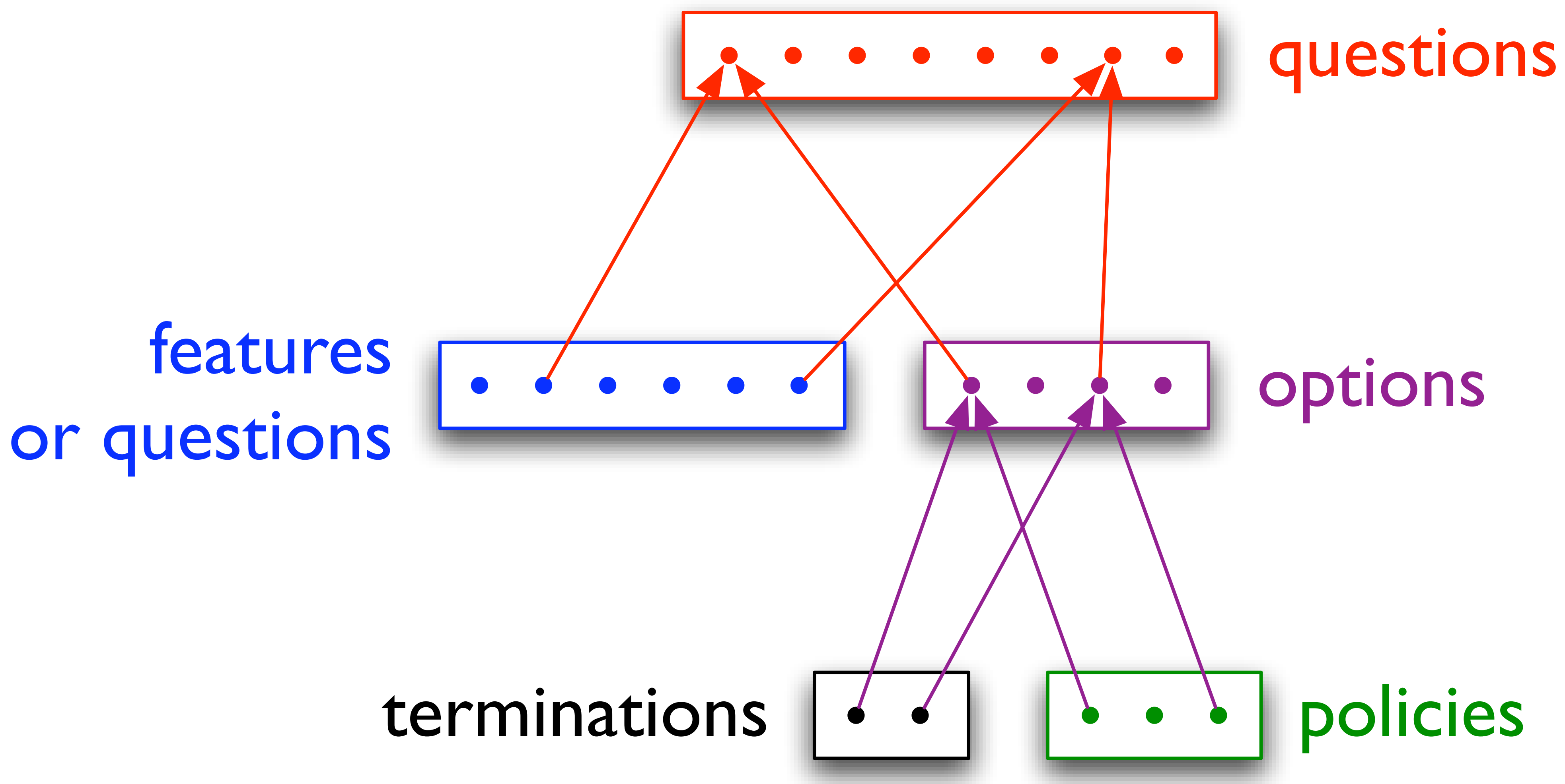


Step Forward
Turn Left
Turn Right
Wander 20 steps
Time step = 200003
L
F
R
Obs
Wander
Leap
L
R
Leap
Leap

Conclusions from demonstration

- The TD network learned much of the commonsense knowledge of the micro-world
- The world is highly non-Markov – the TD net maintained substantial short-term memory
- Large-scale knowledge can be learned even when short-term cannot
- Micro-worlds can be used to effectively illustrate ideas and test algorithms

Question network structure



Question networks define the semantics of the predictions

Learning in TD Networks

- Think of each option as a kind of demon, examining the actions and observations as they flow by, in the context of the current state
- If the action is inconsistent with an option's policy, then its questions don't learn
- If the action is consistent, then learning will occur
  - the observation is examined to see if the option has terminated (completed)
  - if it has, then all predictions about it are incremented toward the value of their target signal
  - if hasn't, then a TD update is done: all predictions are incremented toward their newly predicted value

To complete the package...

- Need projection, planning (very close)
- Need systematic exploration
- Need off-policy learning
- Need discovery of questions and options

But none of this is required for the main prize:
an AI that can tell for itself whether it is working correctly

Steps toward a predictive AI

1. Representation
2. Verification
3. Learning
4. Planning
5. Exploration
6. Discovery
7. Scaling

Take-home messages

- AI should be oriented around experience
  - but it's not
- Knowledge must be predictions
  - but that's nearly unimaginable
- Predictions can be really complex, abstract, expressive and compositional
  - while their machinery is simple and uniform
- Run-time verification may enable big AI
  - although I will show you just small AI

Thank you for your attention

Key point

- Questions provide *subgoals* for learning
- Enabling useful learning to occur without waiting for reward
- This is the same idea as learning a model of the world's dynamics
- But greatly extended by abstracting in state and time

Pros and cons of predictive grounding of knowledge

- Loses
  - easy expressiveness
  - coherence with people
  - interpretability, explainability
- Gains
  - the knowledge means something to the machine
  - can be mechanically maintained/verified/tuned/learned
  - suitable for general-purpose reasoning methods