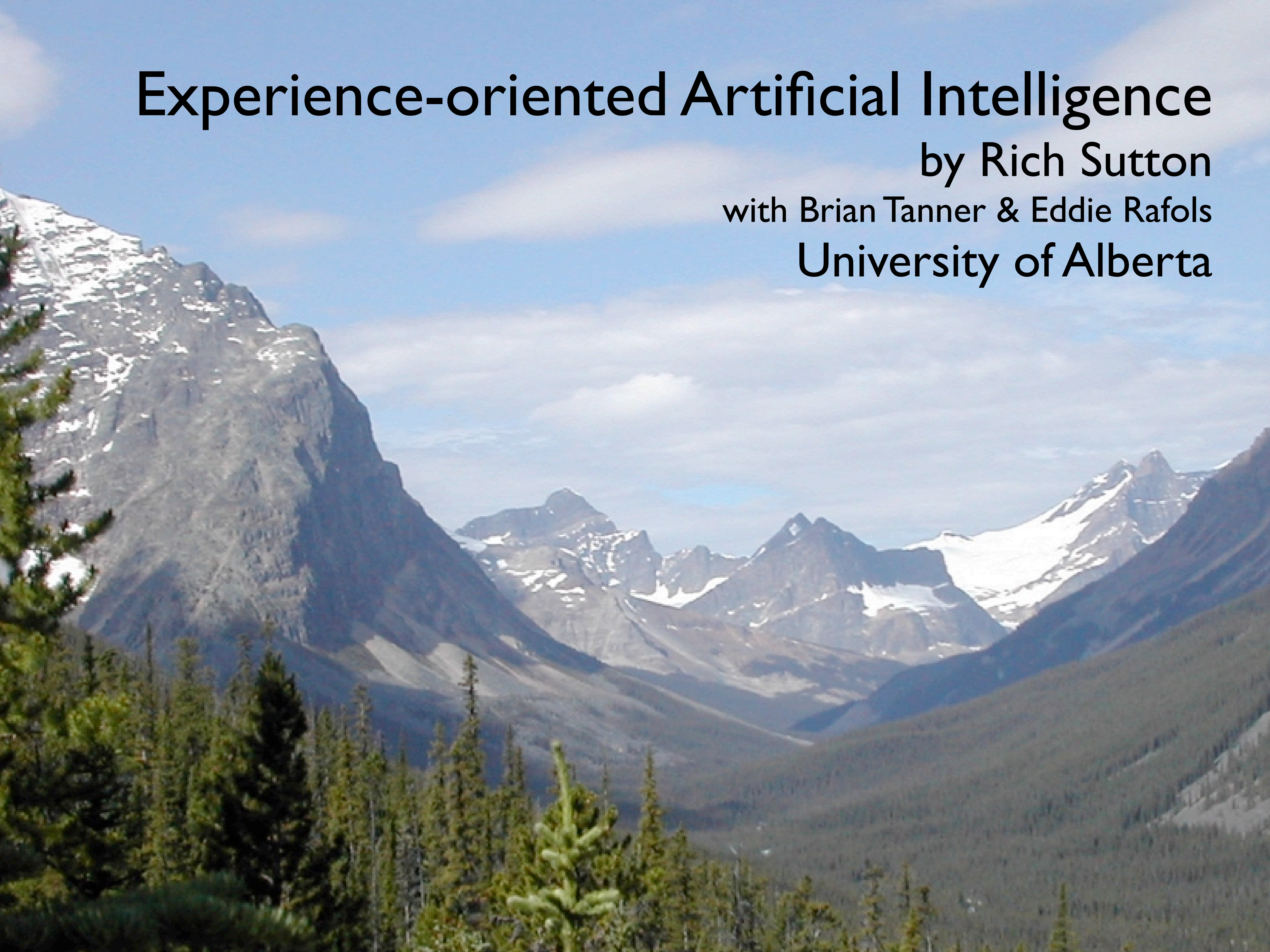
Experience-oriented Artificial Intelligence

by Rich Sutton
with Brian Tanner & Eddie Rafols
University of Alberta



R L & A I

Reinforcement Learning and Artificial Intelligence

PIs:
Rich Sutton
Michael Bowling
Dale Schuurmans
Vadim Bulitko

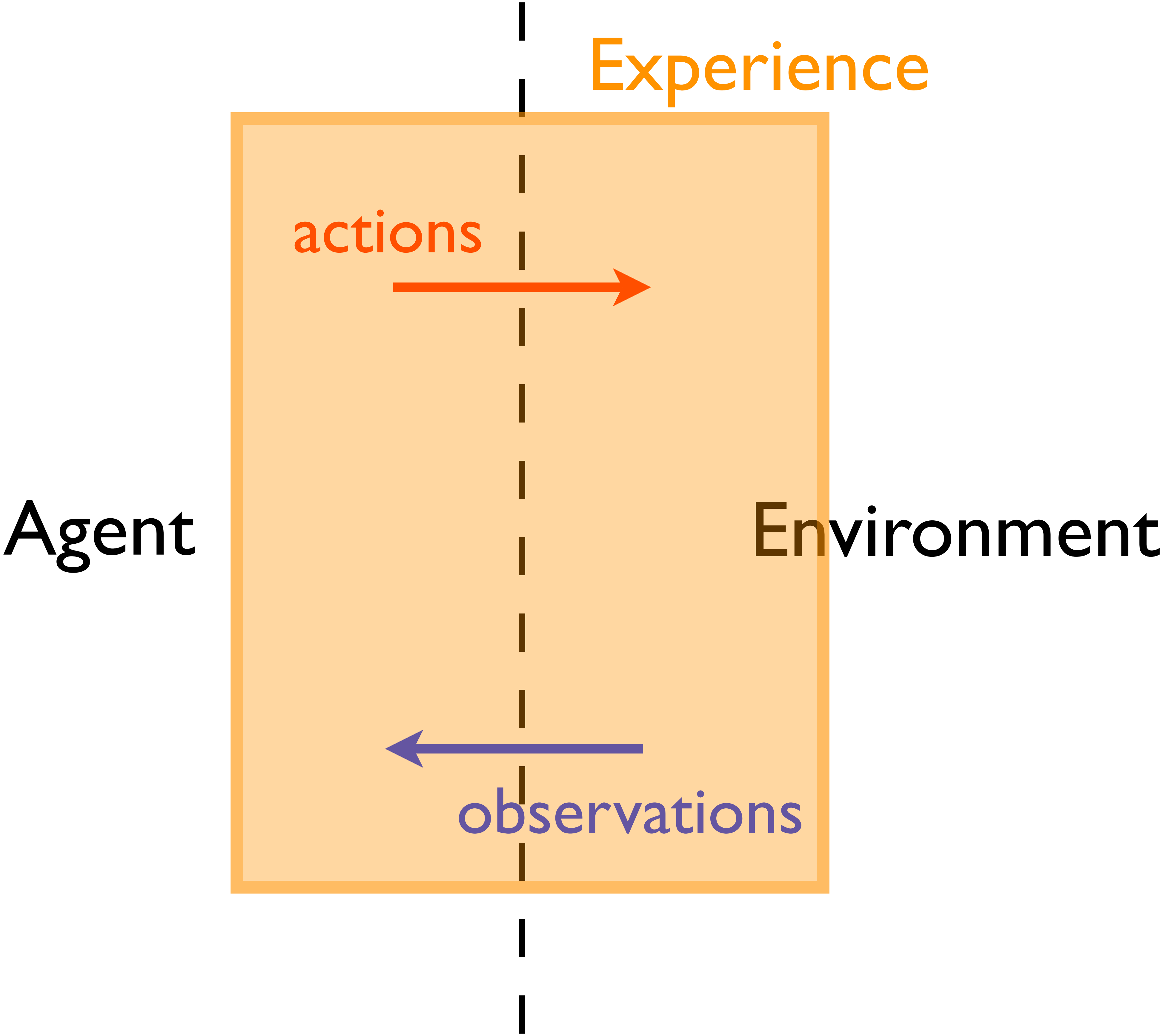
University of Alberta
Quaecumque Vera

Alberta Ingenuity Centre for Machine Learning

Informatics iCORE Circle of Research Excellence

Outline

- Empirical (experience oriented) AI
- A Grand Challenge
- TD networks with temporal abstraction
  - ideas and motivations
  - question and answer networks
  - temporal abstraction
- Illustrative Experiments


Experience
actions
Agent
Environment
observations

Experience is the data of AI

observation

action

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

future

past

now

The special thing about life is that it has a now

Experience is what life is all about

Experience is the final common path,
the only result of all that goes on
in the agent and world

Experience matters computationally

- Experience is the most prominent feature of the computational problem we call AI
- It's the central data structure, revealed and chosen over time
- It has a definite temporal structure
  - Order is important
  - Speed of decision is important
- There is a continuous flow of long duration (a lifetime!)
  - not a sequence of isolated interactions, whose order is irrelevant

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 "respect" 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

Yet we rarely see these basic AI issues discussed
in terms of experience

Is it possible or plausible that they could be? Yes!

Would it matter if they were? Yes!

Outline

- Empirical (experience oriented) AI
- A Grand Challenge
- TD networks with temporal abstraction
  - ideas and motivations
  - question and answer networks
  - temporal abstraction
- Illustrative Experiments

Warning:
big issues,
small experiments

The Problem

- How can we represent complex, commonsense knowledge of the world?
- With mathematical clarity
  - With meaning is as clear as that of a transition probability
- In such a way that it is maintainable without continuous human intervention
- In such a way that it can be learned and used flexibly (e.g., for planning)

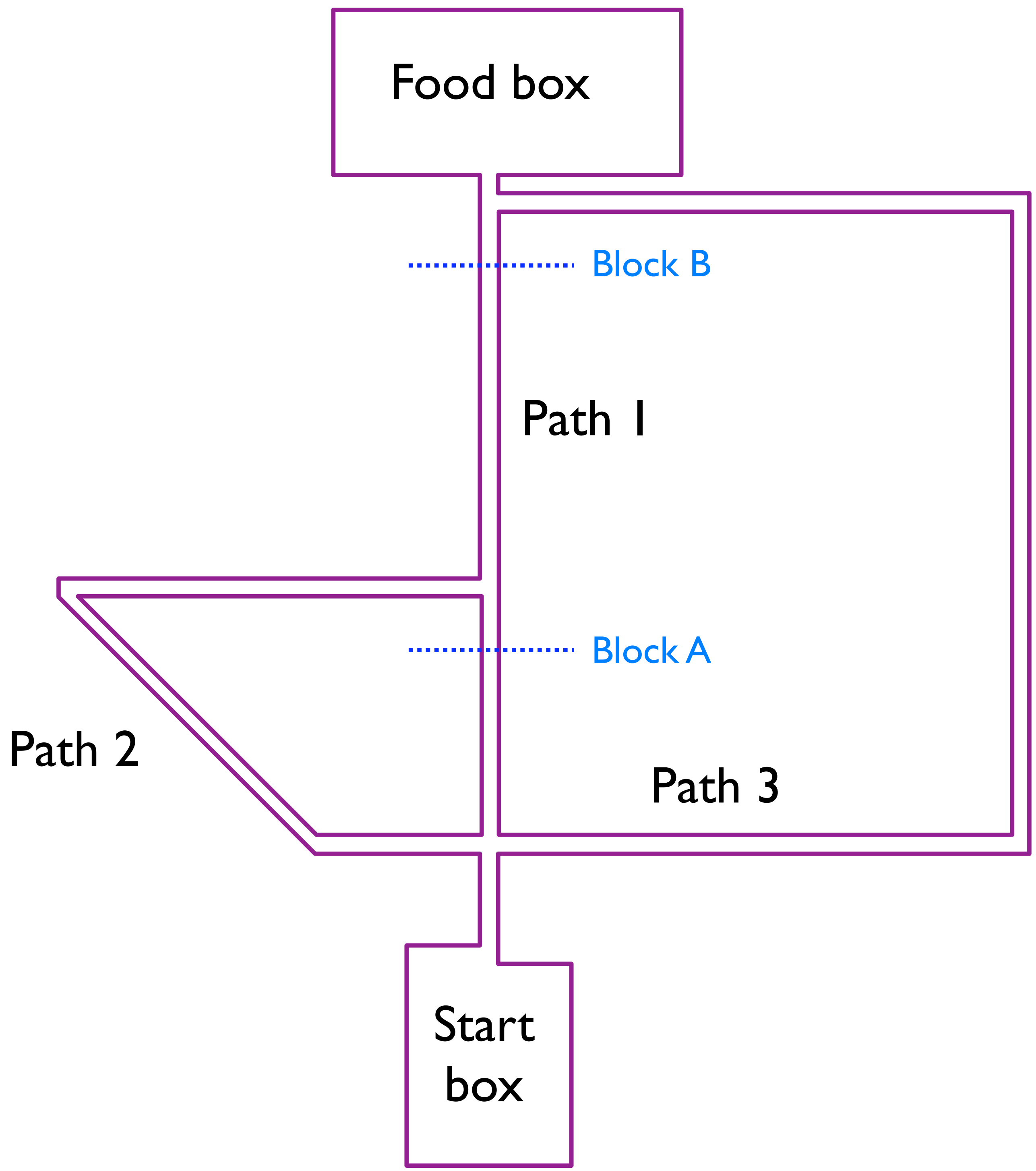
A new empiricism

- Knowledge is about experience
  - knowledge is prediction of experience
  - enables verification, self-maintenance
- But not necessarily from experience
  - takes no position on the nature/nurture debate
- Does not require public agreement
  - specific to agent's sensory and motor capabilities
  - subjective is primary; objective secondary

Philosophical and Psychological Roots

- Like classical british empiricism (1650–1800)
  - Knowledge is about experience
  - Experience is central
- But not anti-nativist, not *tabula rasa*
- Emphasizing sequential rather than simultaneous events
  - Replace association/contiguity with prediction/contingency
- Close to Tolman's "Expectancy Theory" (1932–1950)
  - Cognitive maps, vicarious trial and error
- Psychology struggled to make it a science (1890–1950)
  - Behaviorism, operational definitions
  - Objectivity

Tolman & Honzik, 1930
“Reasoning in Rats”



Subjective Empiricism Hypothesis

Everything we know that is specific to this world
(as opposed to universally true in any world)
is a prediction or memory of experience

A Grand Challenge:
Grounding knowledge in experience

- To represent human-level world knowledge solely in terms of experience:
  - observations
  - actions
  - time
- Why? So that the knowledge can be maintained, refined, learned, and used *autonomously*

Outline

- Empirical (experience oriented) AI
- A Grand Challenge
- TD networks with temporal abstraction
  - ideas and motivations
  - question and answer networks
  - temporal abstraction
- Illustrative Experiments

Goals for TD nets

- Represent commonsense knowledge
  - break AI/RL problem into subproblems
- Model dynamical systems
  - with abstraction of state (PSRs, TD nets)
  - with abstraction of time (options)
- Predictive knowledge
- Subjunctive knowledge - “if I were to...”
- Compositional knowledge - preds of preds

Examples

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?

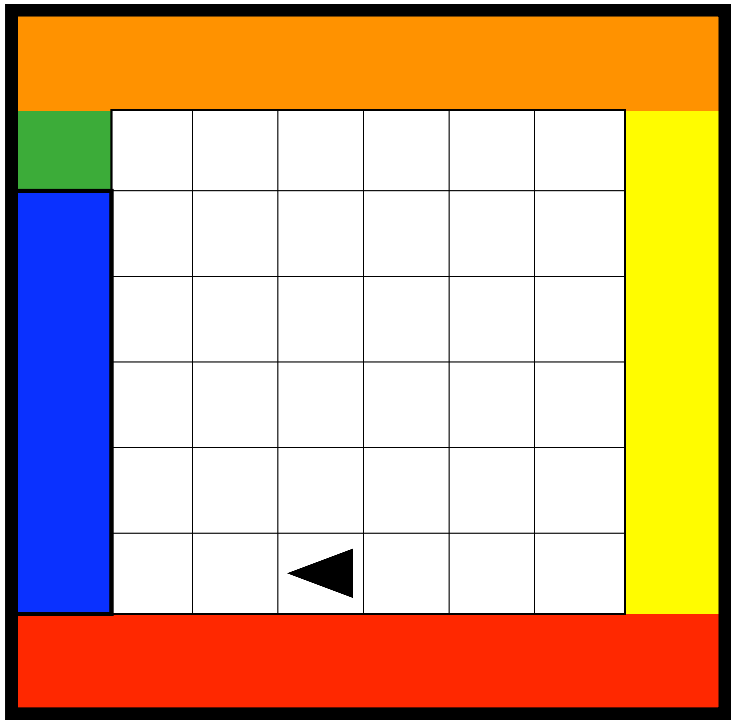
...turn over the glass,
would the carpet be wet?

Outcomes are not primitive observations

They are sets of predictions

Examples in compass world

If I were to...



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

"facing a blue wall"

not compositional

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

compositional

Novelties vis-a-vis PSRs

- Conventional PSRs predict *low-level* observations conditional on *low-level* actions
  - not compositional
  - not temporally abstract

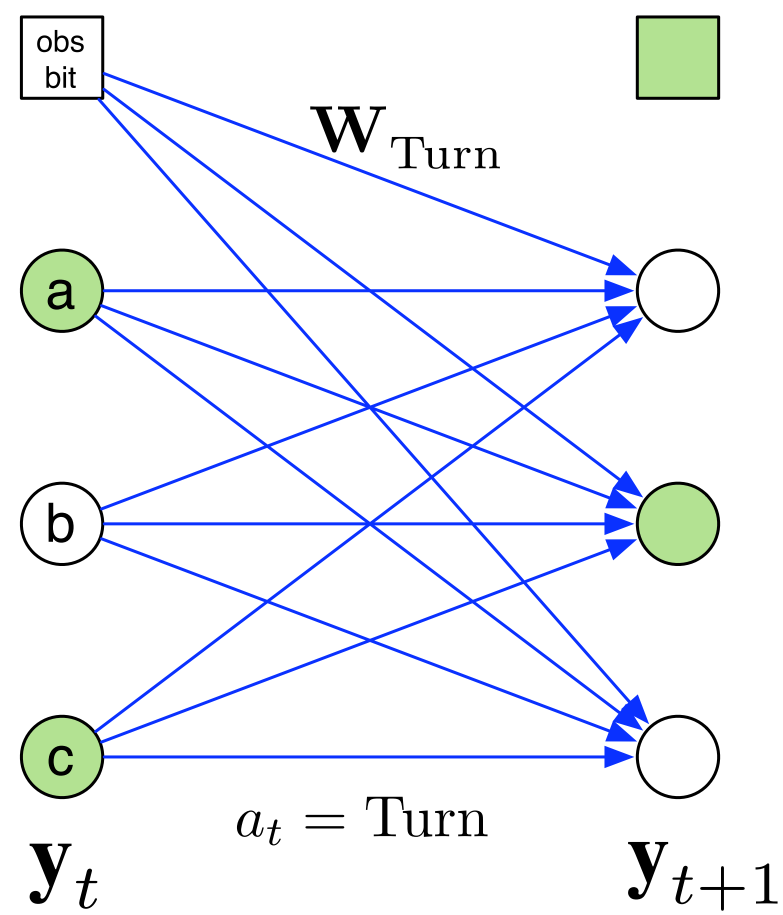
Novelties vis-a-vis options

- State abstraction
- Function approximation
- No Markov assumption
- Policies → recognizers
  - accept a set of actions

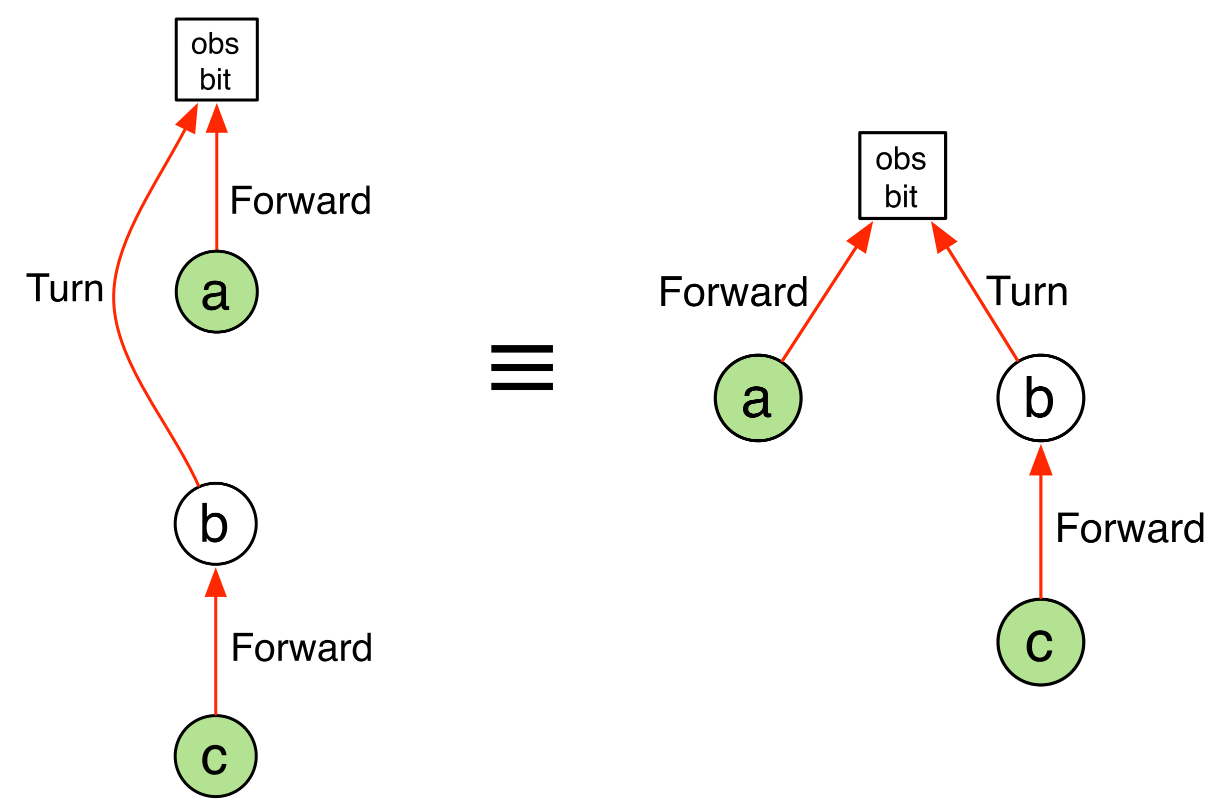
Temporal-difference networks

- 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

Answer Network
(computes predictions)



Question Network
(affects learning)



e.g., $\mathbf{y}_{t+1} = \boldsymbol{\sigma}\left(\mathbf{W}_{\mathrm{Turn}}^{T}\phi(\mathbf{y}_t)\right)$

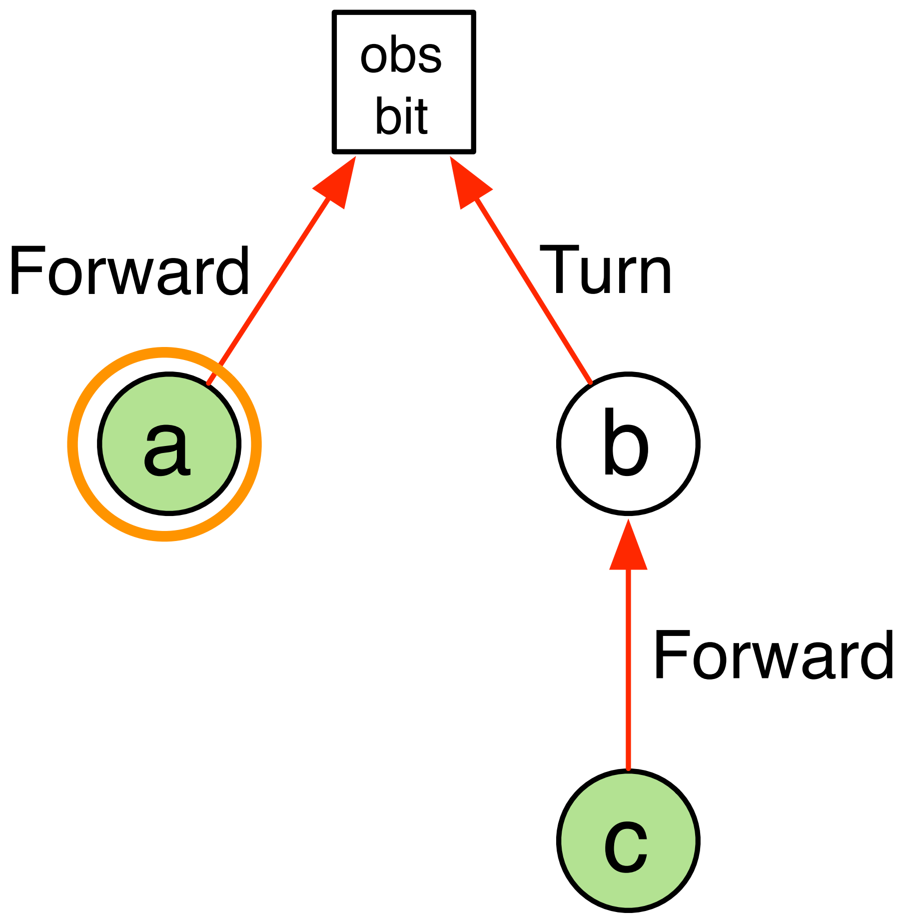
matrix for Turn

feature vector

$$\mathbf{y}_t \xrightarrow{\alpha} \mathbf{z}_t$$

target = the thing pointed at

Question network
(affects learning)



Node a = if I were to step Forward, what would the observed bit be?

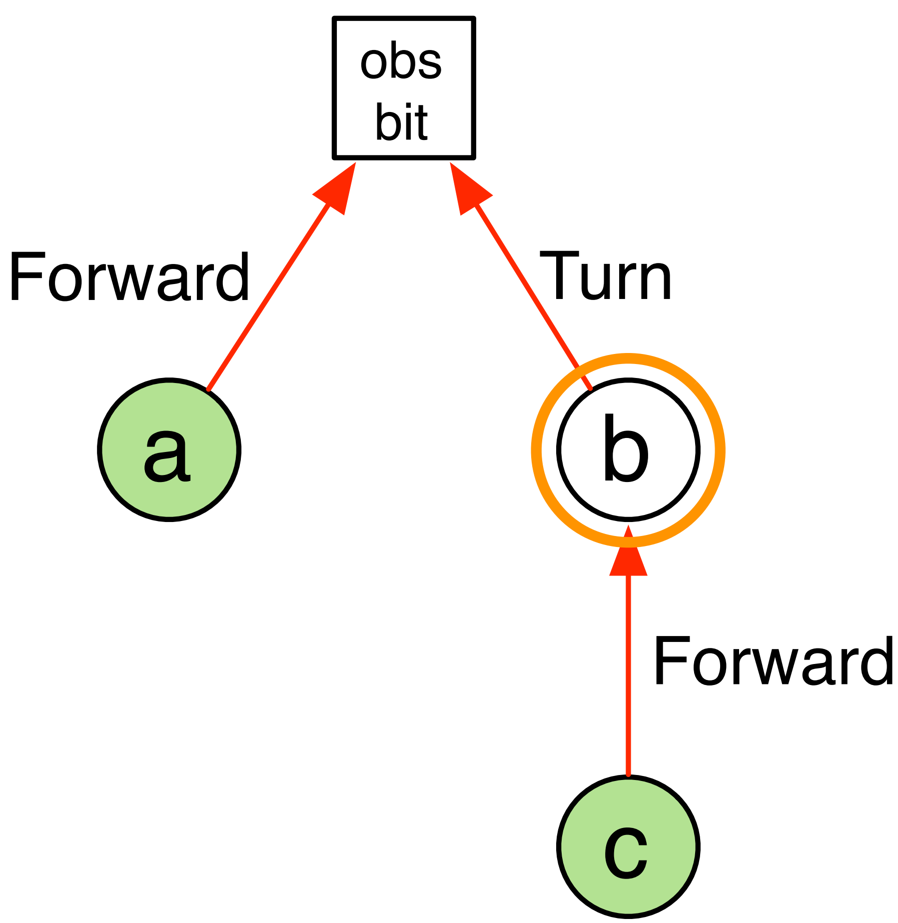
$$z_t^a = o_{t+1} \qquad c_t^a = 1 \text{ iff Forward}$$

target

condition

observed bit

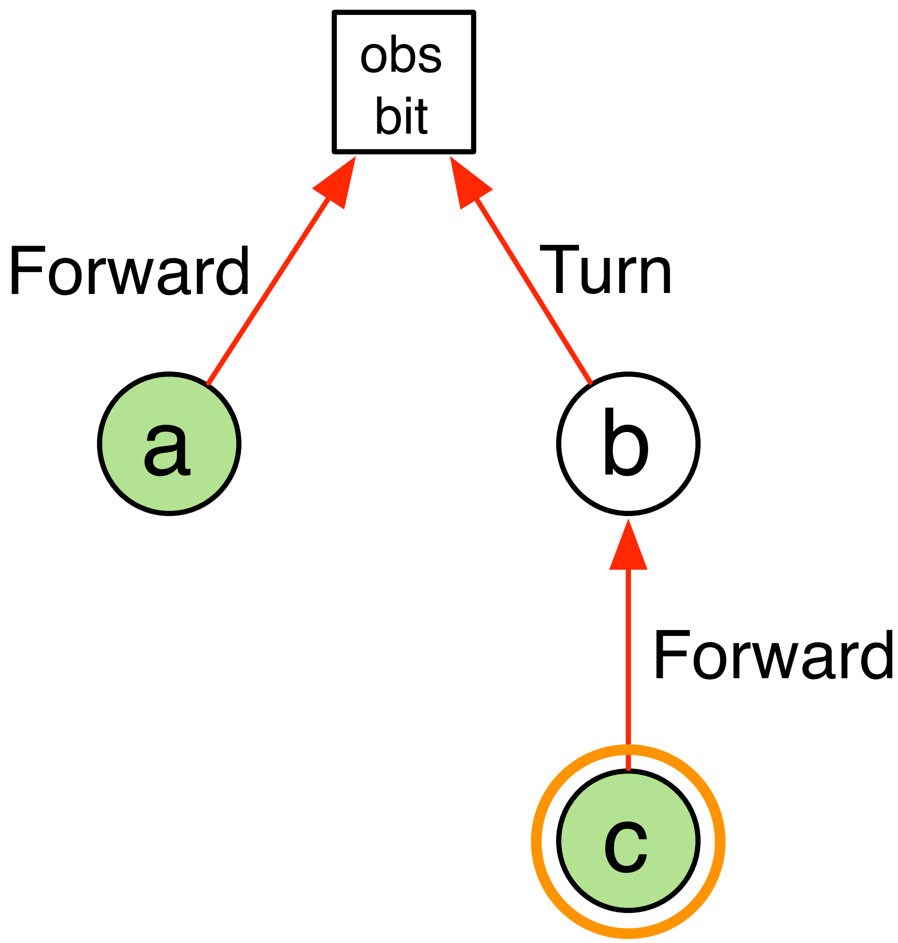
Question network
(affects learning)



Node b = if I were to Turn,
what would the observed bit be?

$$z_t^b = o_{t+1} \qquad c_t^b = 1 \text{ iff Turn}$$

Question network
(affects learning)



Node c = if I were to step Forward, what would the *prediction* for Turn be?

$$z_t^c = y_{t+1}^2 \qquad c_t^c = 1 \text{ iff Forward}$$

Not: if I were to step Forward, then Turn, what would the observed bit be?

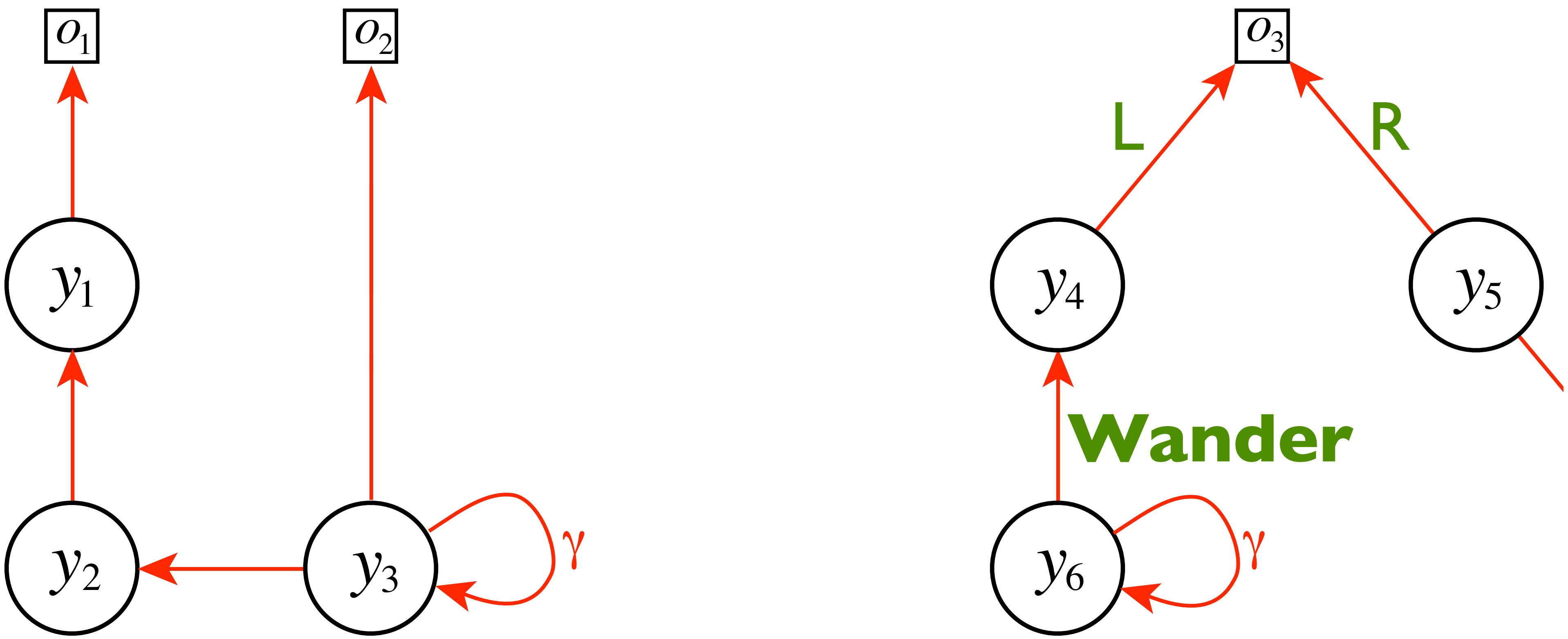
TD Semantics

- The inter-predictive relationships in the question network

Extensive Semantics

- The inter-predictive relationships unrolled until they ground out in observations and actions

More question networks

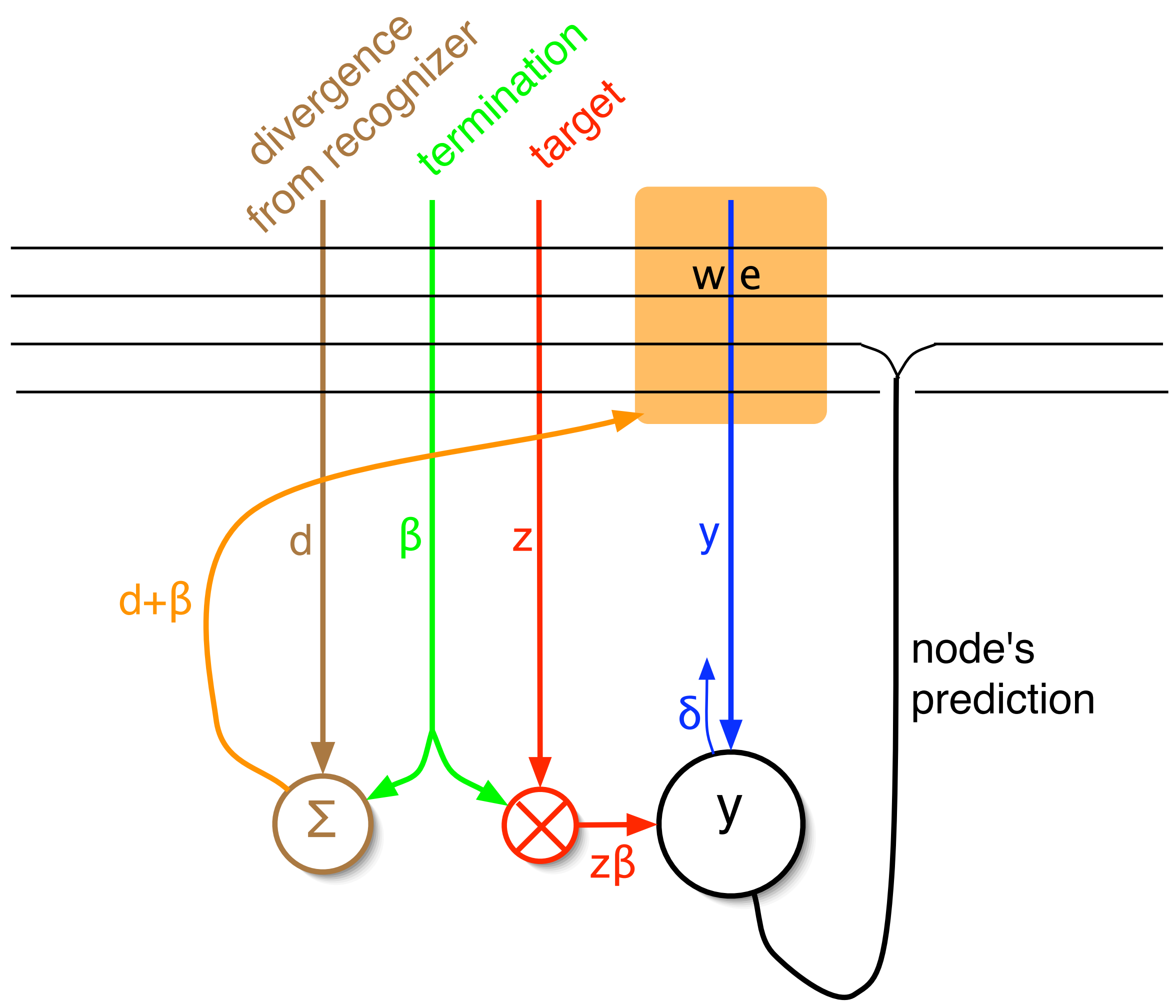


Learning in TD Networks

- Predictions from the answer net are compared to targets from the question net
- Weights in the answer net are changed to make future predictions closer to the targets

Learning in TD networks, step by step

1. Let y_i be the prediction of node i, i=1,...,n
2. Take action *a*
3. Set $c_i \in \{0,1\}$ – does *a* satisfy question i's condition?
4. Observe *o*
5. Create feature vector **x** characterizing **y**,*a*,*o*
6. Construct new prediction vector $\mathbf{y}' = \mathbf{W}\mathbf{x}$ (answer net)
7. Construct targets z_i for each y_i, i=1,...,n (question net)
8. Construct TD errors $\delta_i = z_i - y_i$
9. Update weight matrix $\Delta w_{ij} = \alpha \, c_i \, \delta_i \, x_j$



Circuit diagram for the TD net learning algorithm

Continuous-time equations

$$\dot{y} = u(a, o, y, \vec{w})$$ answer net (pred. update)

$$\dot{w}_i = \alpha\, \delta\, e_i$$ learning (weight update)

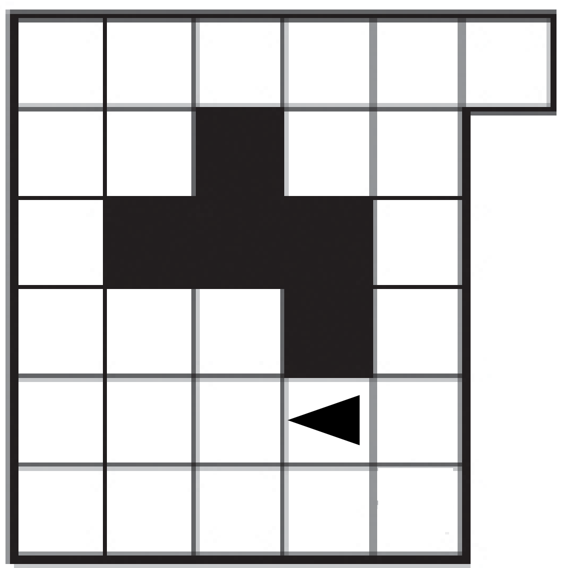
$$\delta = \dot{y} + z \cdot \beta$$ TD error

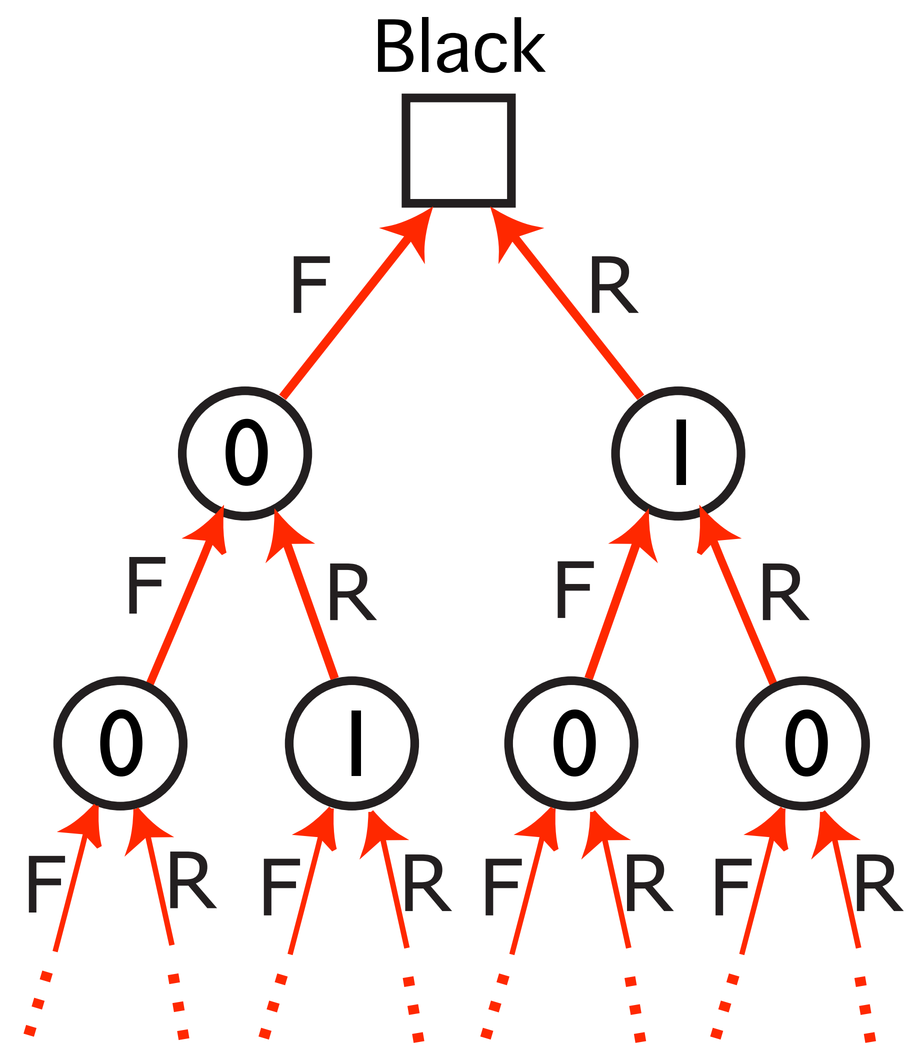
$$\dot{e}_i = \frac{\partial \dot{y}}{\partial w_i} - (d + \beta)\, e_i$$ eligibility traces

Outline

- Empirical (experience oriented) AI
- A Grand Challenge
- TD networks with temporal abstraction
  - ideas and motivations
  - question and answer networks
  - temporal abstraction
- Illustrative Experiments

Bit-to-bit gridworld

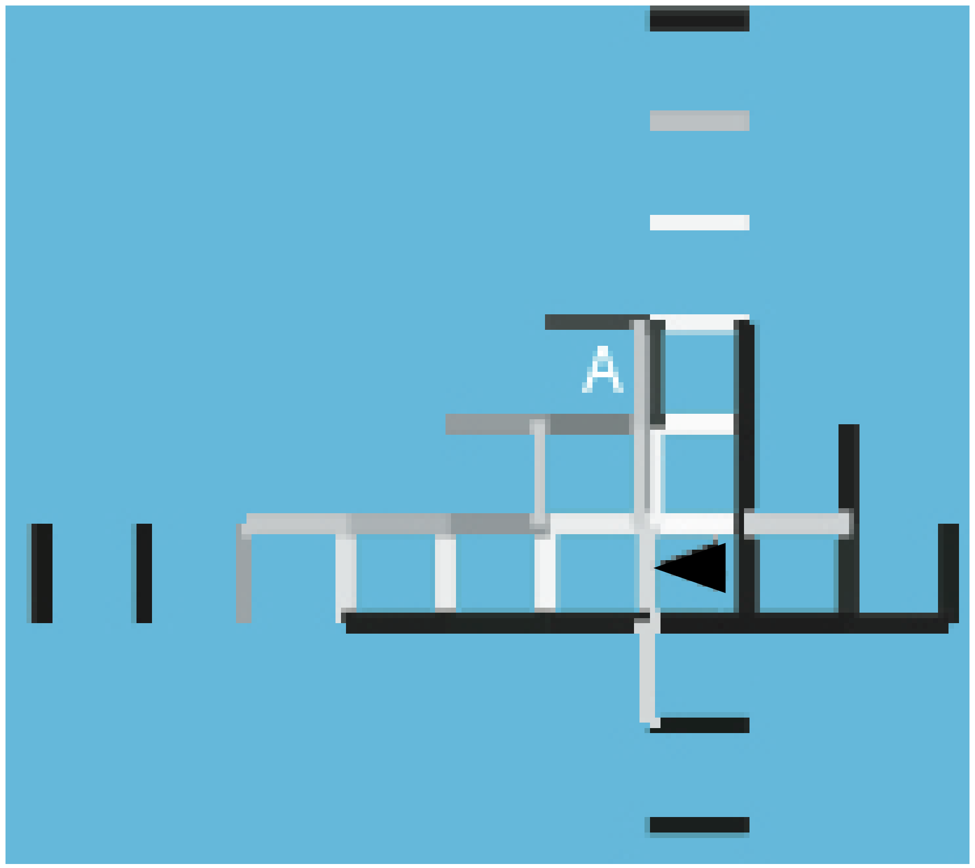


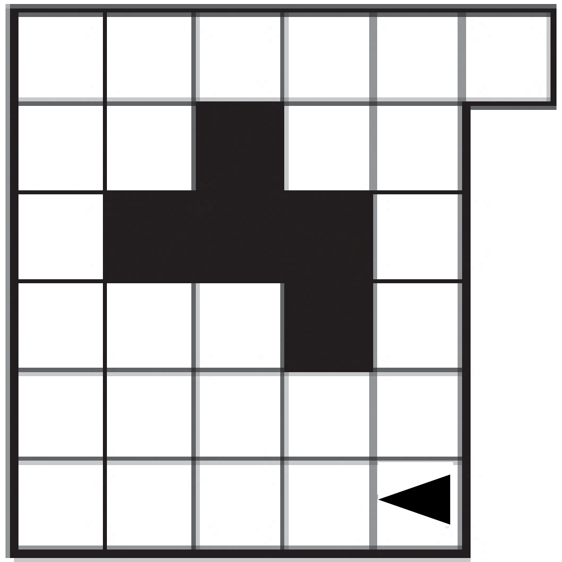


sensations: black/white ahead
actions: F(orward), R(ight)

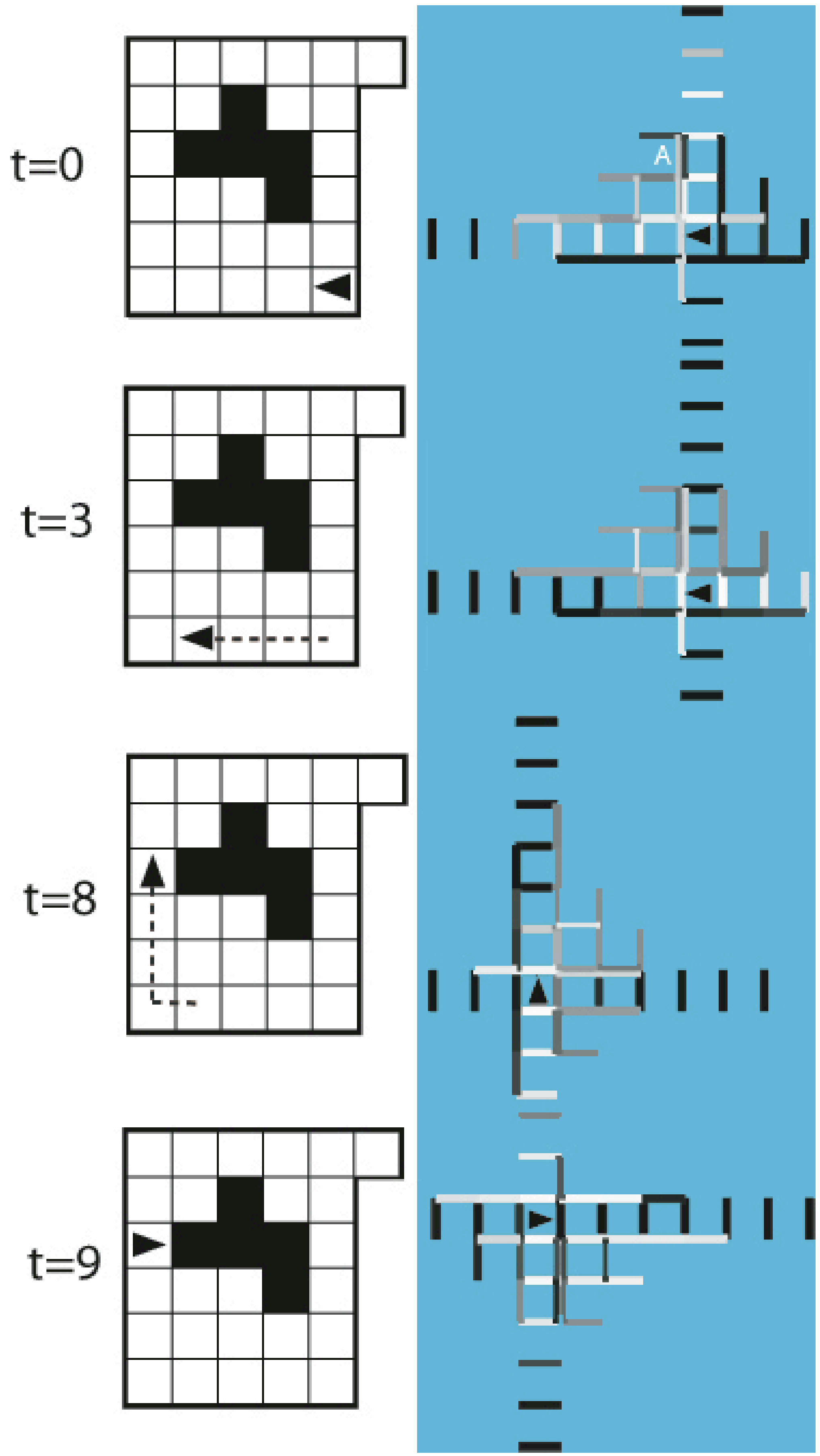
one bit each of sensation and action!

Subjective display





- Each line represents the answer to one of the agent's questions
- Lines are colored to indicate the probability that a certain sequence will lead to an observation
  - Black is very probable
  - White is unlikely
  - Shades between represent a measure of uncertainty



- knows wall on left & behind
- knows wall ahead 4 or 5 steps
- knows some walls around ‘A’

- knows wall ahead closer
- knows no wall behind
- knows wall on right

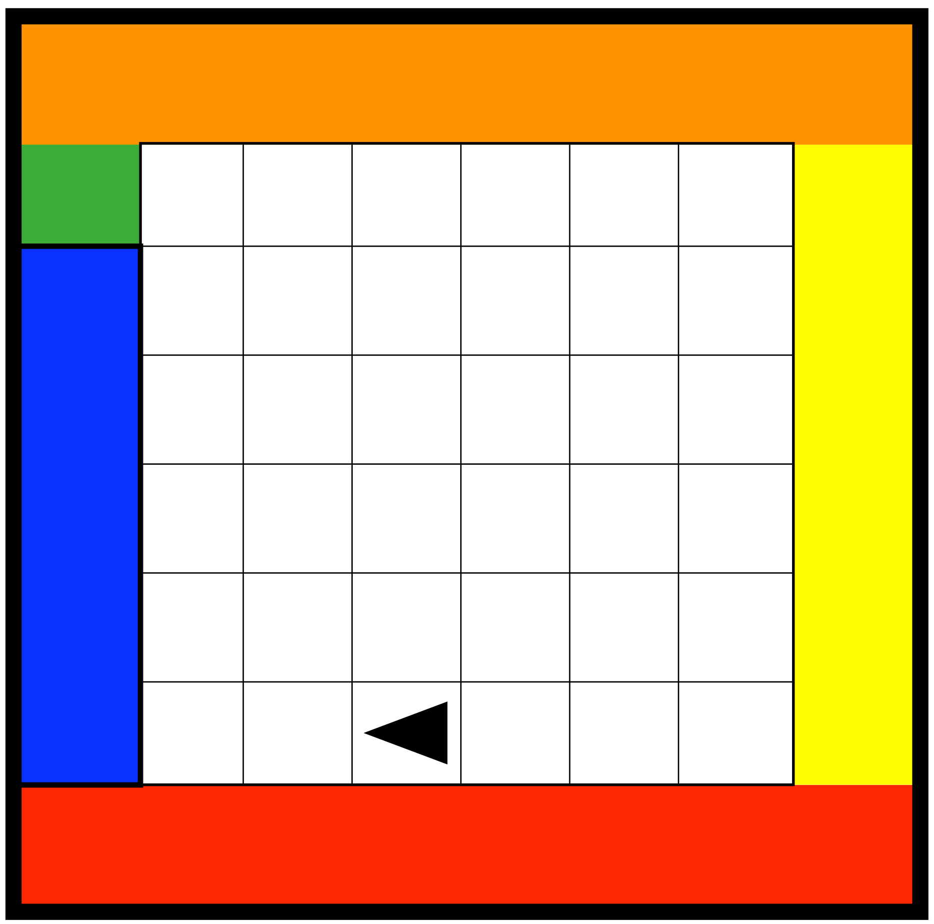
- knows wall ahead closer
- unsure about center object

- knows blocked
- most of rest not known

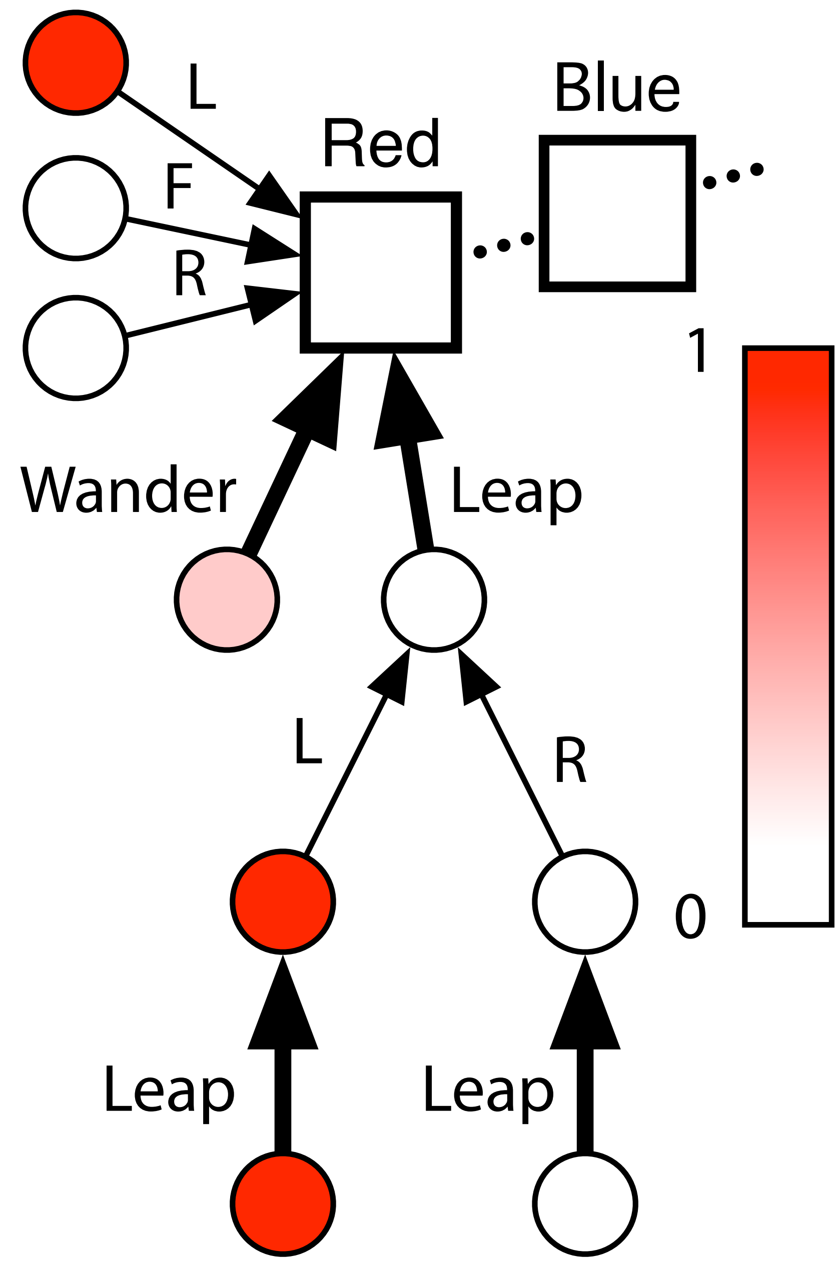
by inspection, also knows

- that if it is facing a wall and goes forward, then it will always see a wall
- that going forward when facing a wall does not change any predictions
- that four consecutive turns does not change any predictions
- basic rotational persistence - if shown a wall then walked away, rotated a few times, and walked back, still knows wall is there

World

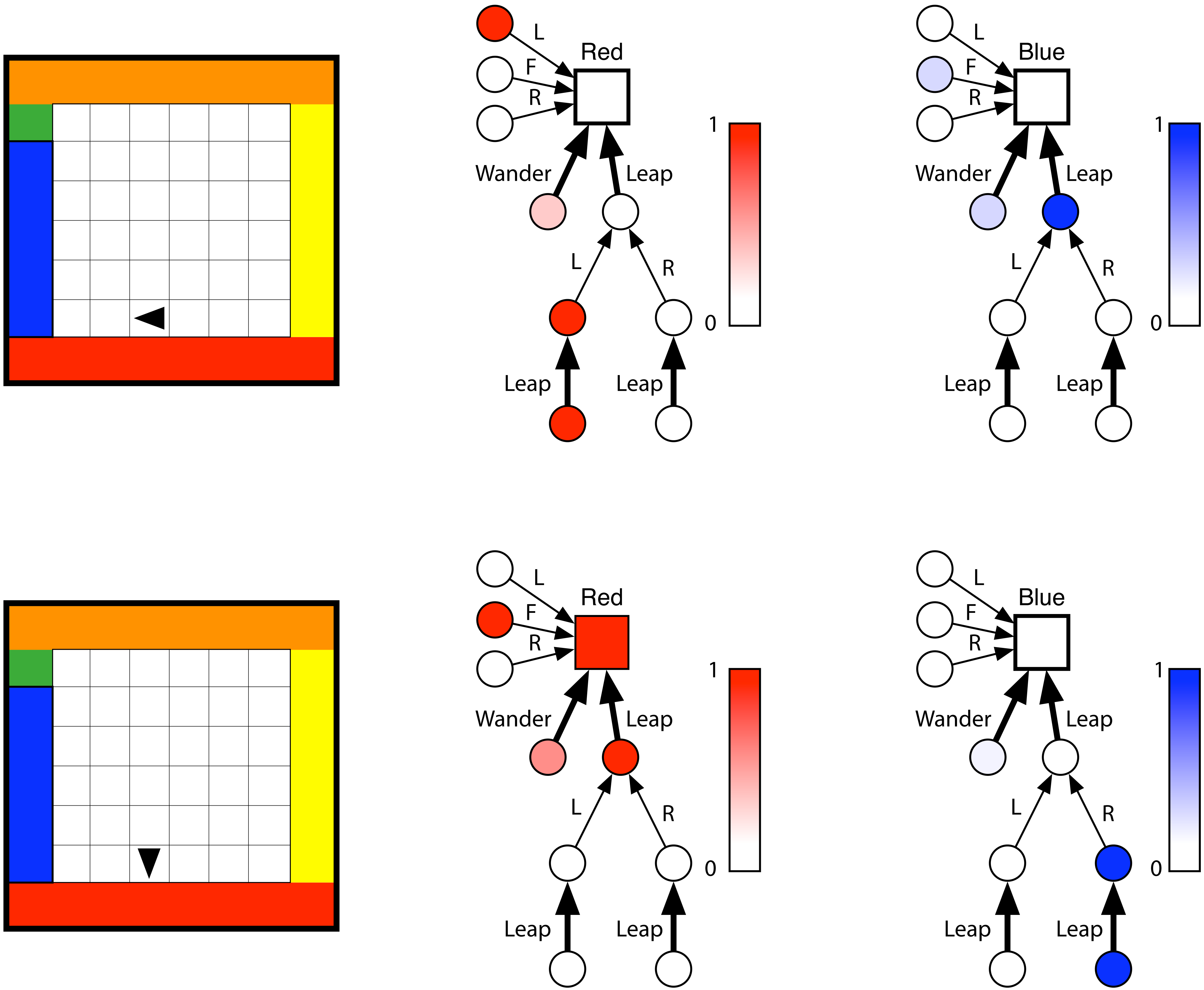


Question net



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

Results after extensive experience wandering randomly



Conclusions

- The TD network learned much of the commonsense knowledge of these micro-worlds (though not all)
- The worlds are highly non-Markov – TD net maintained substantial short-term memory
- Temporally abstract knowledge can be learned, even when short-term cannot
- Micro-worlds can be used to effectively illustrate ideas and test algorithms

Summary

- Subjective experience is the data of AI
- Subjective empiricism is appealing
  - verifiable knowledge
  - explicit, machine-readable semantics
- Explicit representation of questions is necessary
- Abstraction is key – in state and time
  - with compositionality (TD nets)
- Sensori-motor knowledge is rich and complex, and the basis for higher level concepts – connecting the two is an awesome challenge

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

Thank you for your attention