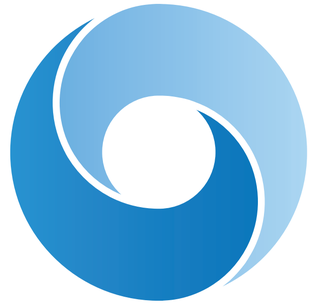
Eyes on the Prize

Rich Sutton
Alberta Machine Intelligence Institute
University of Alberta
DeepMind Alberta
Reinforcement Learning and Artificial Intelligence Lab



20 years ago, Alberta had the foresight to invest $2M/year in machine learning

- which has paid off handsomely in science and commerce in Alberta

What should Alberta invest in today that will pay off in the next 20 years?

- My answer is to double down on the science of AI
- There is still a long way to go; we should keep our Eyes on the Prize of understanding intelligence

Alberta should keep its Eyes on the Prize

- The Prize is to understand the principles of intelligence—what it is and how it works—well enough to create (or become) beings of greater intelligence
- The Prize is a fundamental goal of science, engineering, and the humanities
- Achieving it will change the way we work and play, our sense of self, our sense of life and death, and the goals we set for ourselves and our societies
- Achieving it will be an event comparable in significance to the rise of human life on Earth

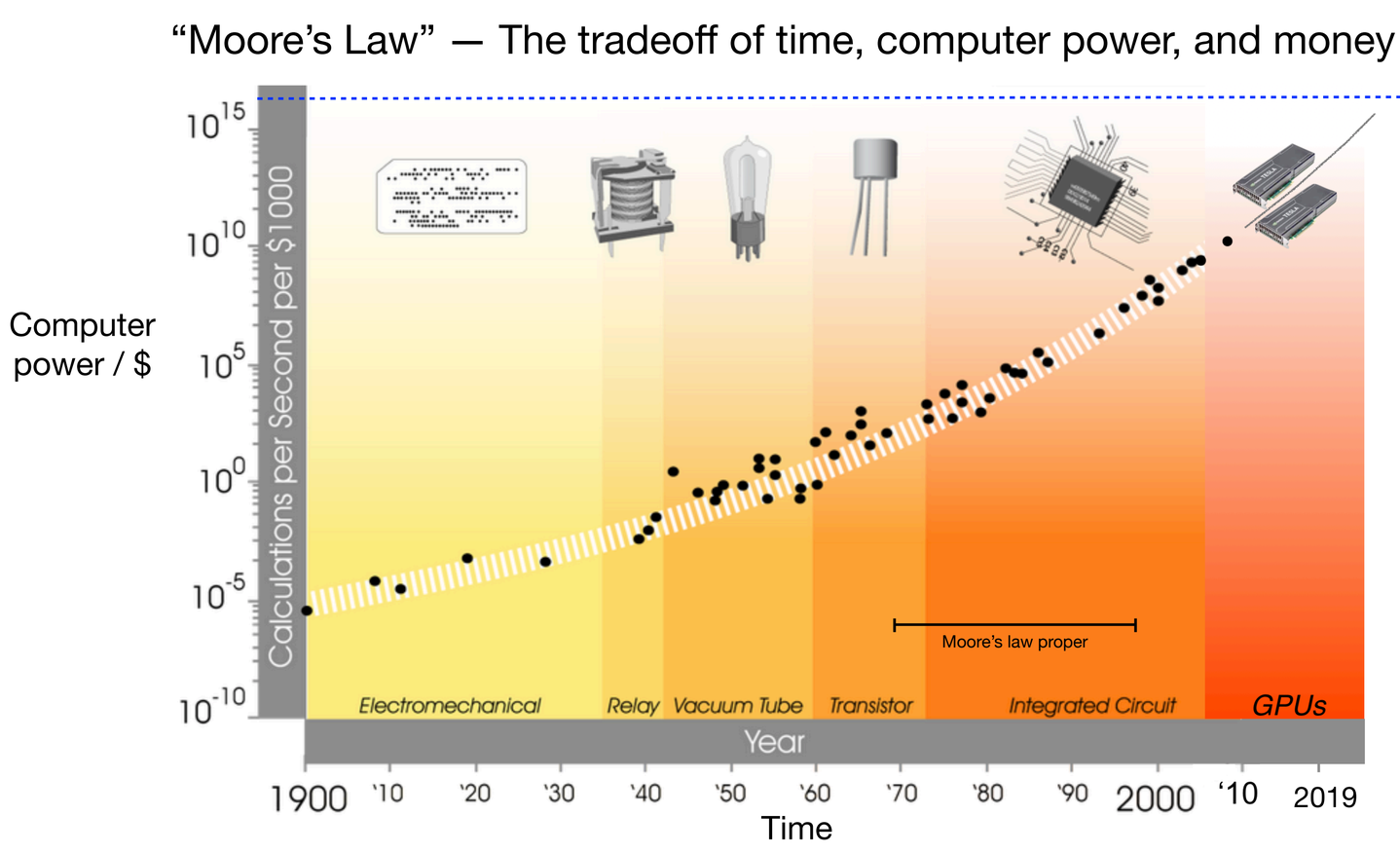
In Alberta, we should strive to be relevant to this great goal

Understanding intelligence would certainly be a big deal

- But some doubt that it is possible
- More doubt that it will happen anytime soon
- Many doubt that it would be a good thing

Is the Prize within reach? And would it be good, or bad?

Computer power/$ increases exponentially, with no end in sight, creating a powerful persistent pressure for understanding intelligence



Brain-scale computer power will cost ≈$1000 in ≈2030

This estimate is rough but robust:
a factor of 10 ≅ 5 years

⇒ AI increases in value
by a factor of 10 every 5 years

And so does the pressure
to find the algorithms/software

I estimate a 50% probability
of human-level AI by 2040

Investing in AI science is a good bet for Alberta

- Human-level AI will probably happen in the next decades;
  in 20 years it is a coin flip

- Even if we lose the coin flip, investing in AI is likely to have been good;
  the partial solutions will likely be commercially valuable

    - Automation and computer power will continue to increase in importance in the economy; Moore's law will chug on

- In 10 years human-level AI is a moonshot; in 20 years it's a coin flip;
  in 50 years it's a near certainty

The Prize is within reach

We don’t get to choose on most big things

- The impact of the printing press and literacy
- The impact of the internet and social media
- The decline of monarchies and rise of democracies
- The shift of world economies from farming to industry
- The shift of Alberta’s economy from oil to technology

Many are impacted. Who are we to choose for all?

Alberta's economy has been dominated by oil, and now it is changing. Is this good or bad?

- This is (or may be) happening. Good for some, bad for others…
- But the real good or bad is how we react to the transformation
  - Do we refuse to accept it and cling to all the old ways?
  - Or do we see, prepare for, and conform to the future?
- So too for the coming of genuine AI
  - We should prepare for and conform to the future
  - We can surf the future, but we can't hold back the wave

Understanding intelligence is inevitable.
We choose to be a part of it because:

- Greater intelligence (better foresight and decisions) in the world means greater productivity with less cost, waste, and conflict

- We will understand ourselves and our place in the universe better, which is what we, as intelligent agents, are built to do

- Selfishly:

    - Early participants in understanding intelligence will reap more of the benefits… and more of the glory

In my opinion, understanding intelligence is one of the few great goods in the universe

Outline

- The Prize of understanding intelligence
- Why we seek it
- How reinforcement learning research in Alberta keeps its eyes on the Prize
    - by focusing on experience

AI research in Alberta (Amii and its eco-system)

- We are doing a mix of applications and longer-term research
    - As Martha White said, we are not yet a leader in the application of RL, but we are striving to be
    - While keeping our eyes on the longer-term Prize of understanding intelligence
- Both are super important, but they are different:
    - To get a successful application, you try to avoid unsolved problems
    - In long-term research, you are drawn to unsolved problems
- Let me tell you a little about what it's like to be drawn to the unsolved problems of AI

First, we characterize the *problem* of intelligence

Intelligence is:

“the most powerful phenomena in the universe” —Ray Kurzweil, 2011

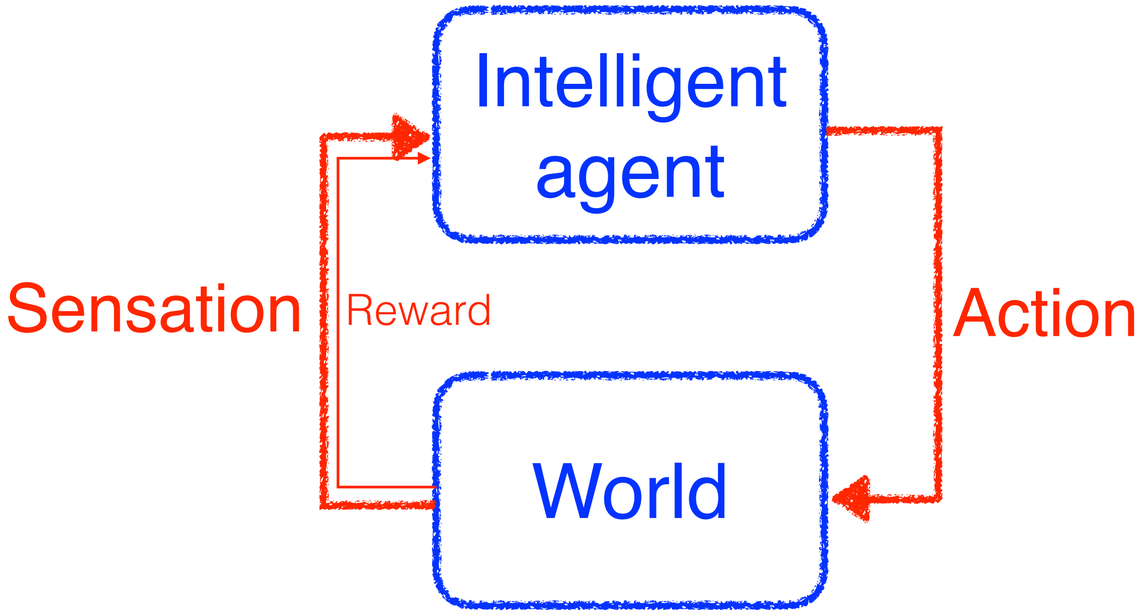
“attaining consistent ends by variable means” —William James, 1890

“the computational part of the ability to achieve goals” —John McCarthy, 1997

“the computational part of an agent’s ability to predict and control a stream of sensations”
—me, now

Intelligence is:

The computational part of an agent's ability to predict and control a stream of sensations, particularly a designated numerical sensation (called reward)

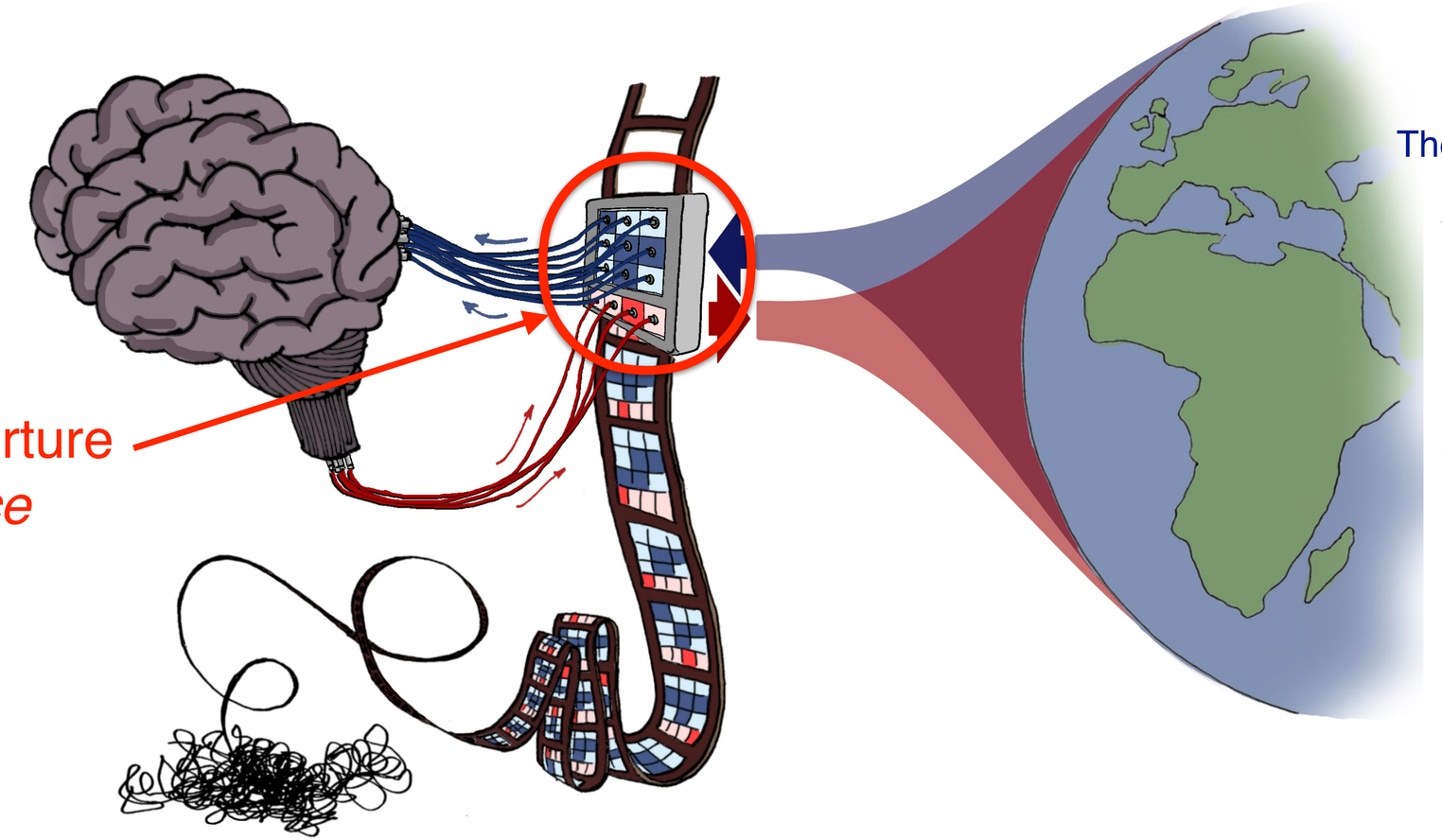


The reward hypothesis:

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

Intelligence is:

The computational part of an agent's ability to predict and control a stream of sensations,
particularly a designated numerical sensation (called reward),
while interacting with a vastly more complex world



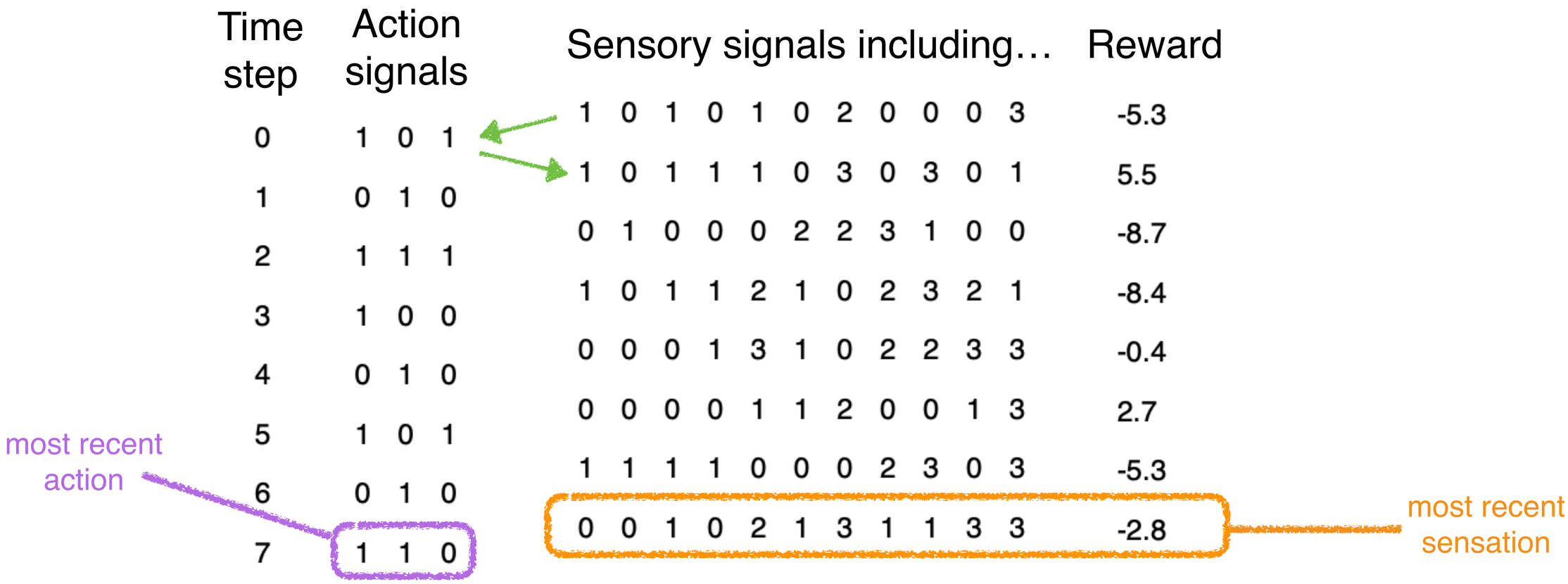
The narrow aperture
of *experience*

The Big-World perspective:

The world is much bigger than the agent

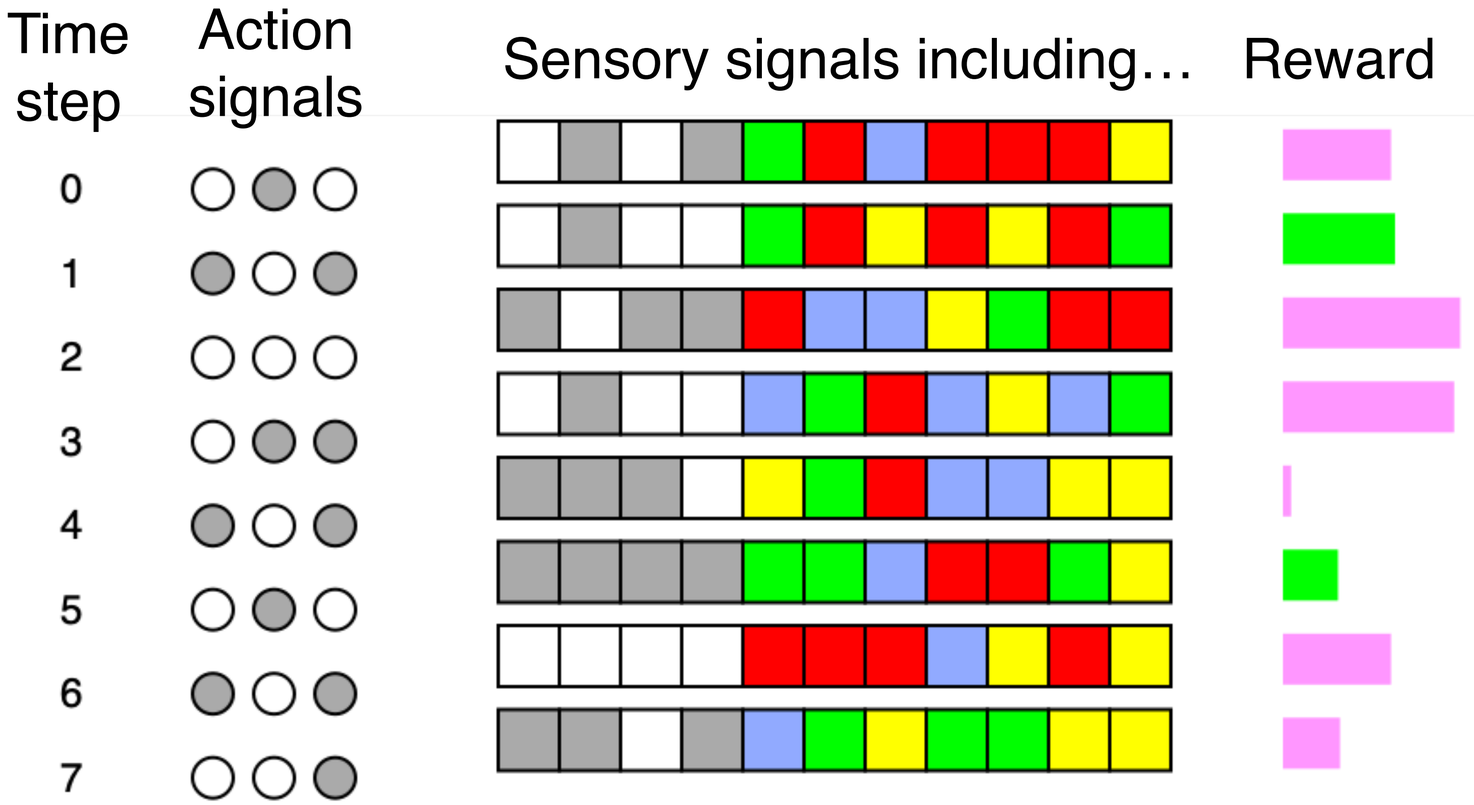
It contains many other agents!

Experience — a concrete nonspecific example



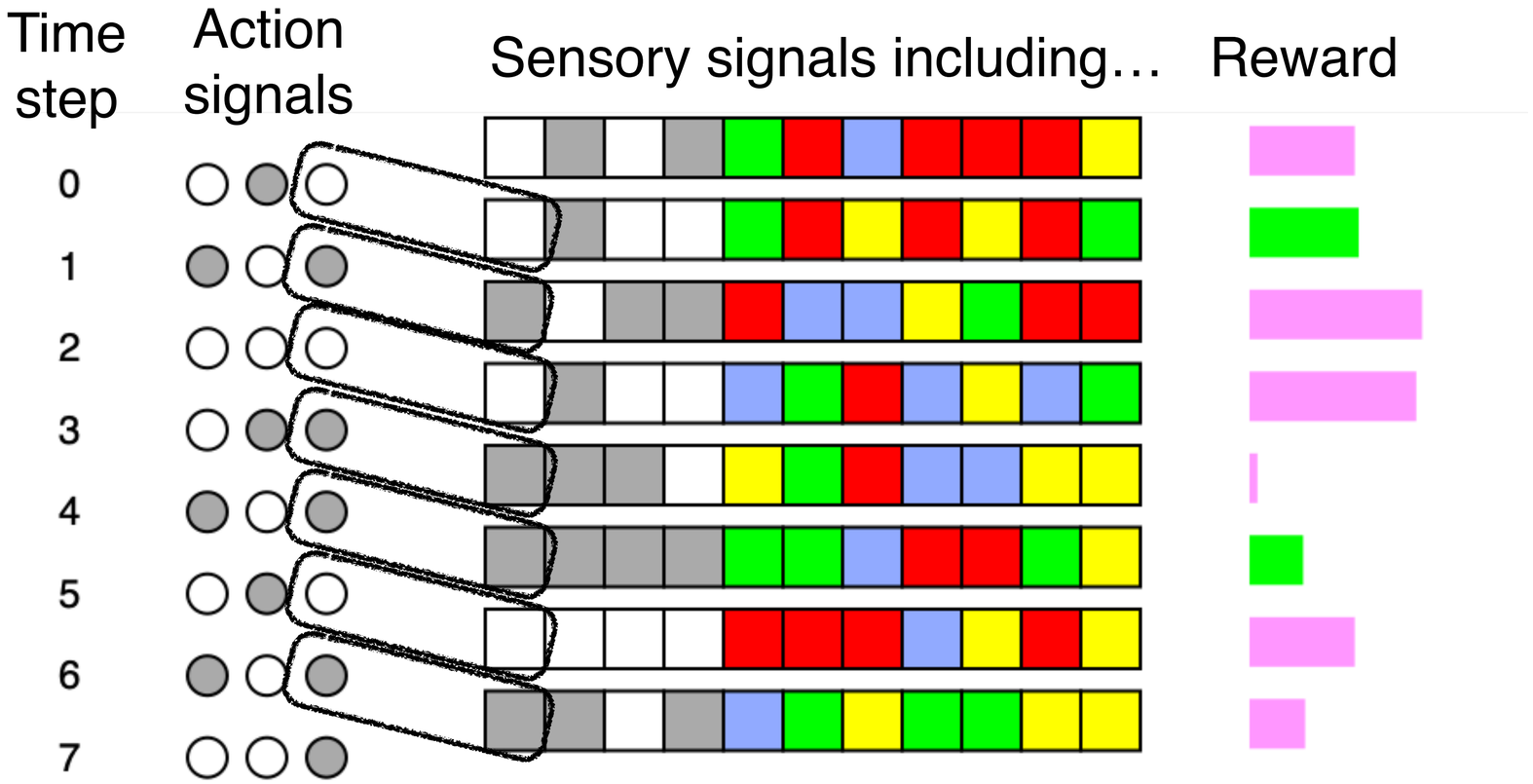
Experience up to time step 7
(think of a time step as ≈0.1 sec)

Experience — a concrete nonspecific example



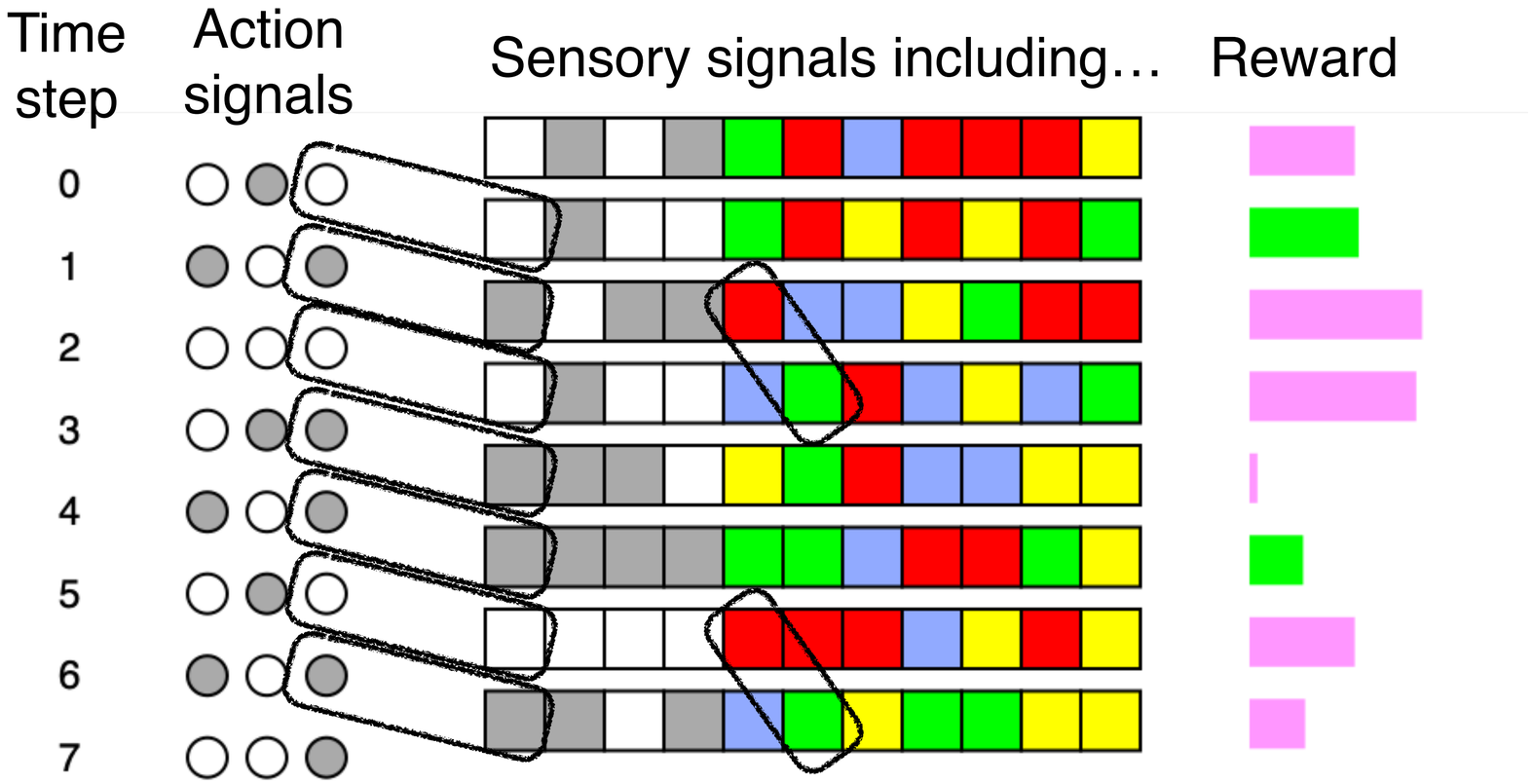
Experience up to time step 7
(think of a time step as ≈0.1 sec)

Experience — a concrete nonspecific example



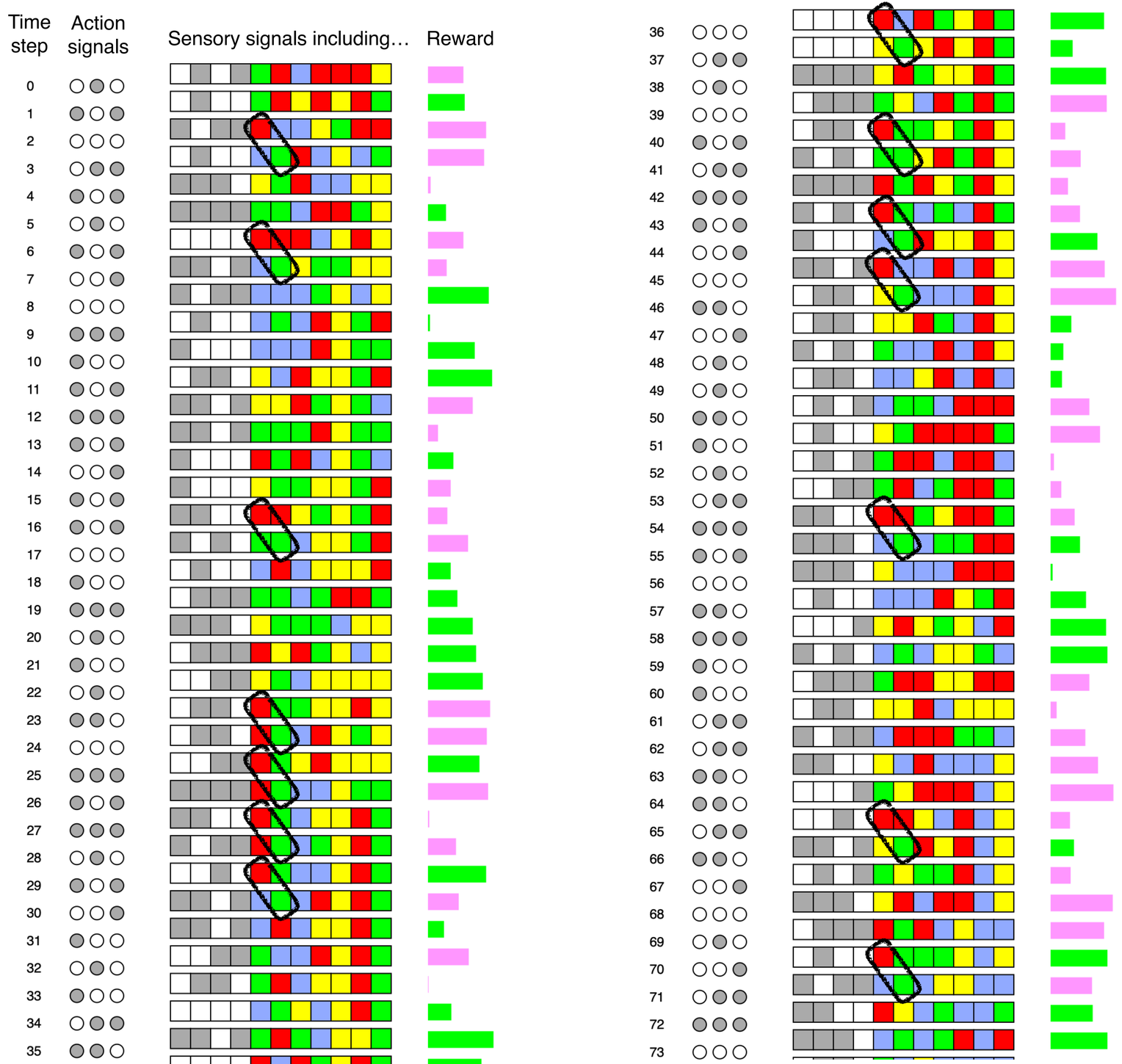
Experience up to time step 7
(think of a time step as ≈0.1 sec)

Experience — a concrete nonspecific example

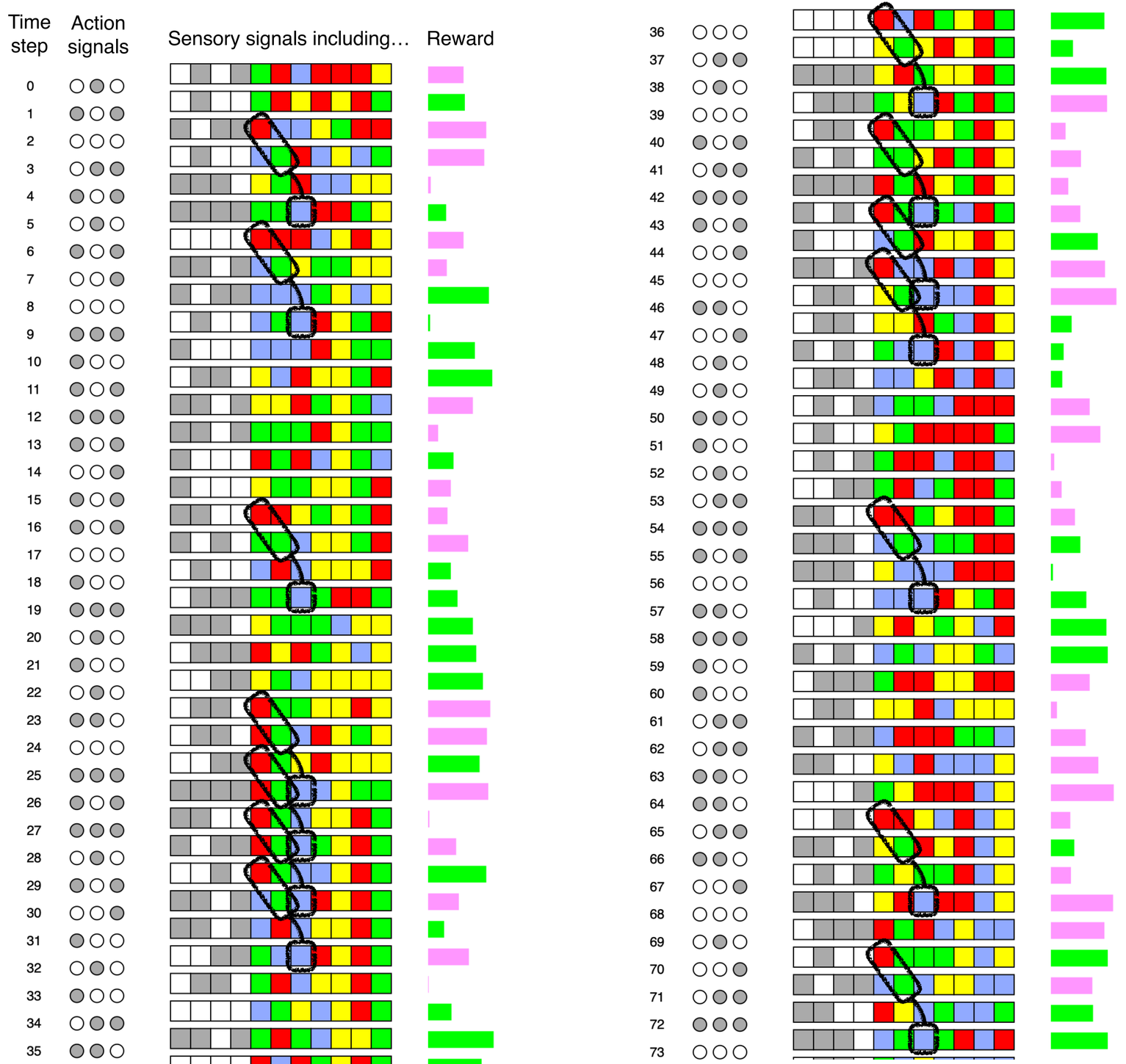


Experience up to time step 7
(think of a time step as ≈0.1 sec)

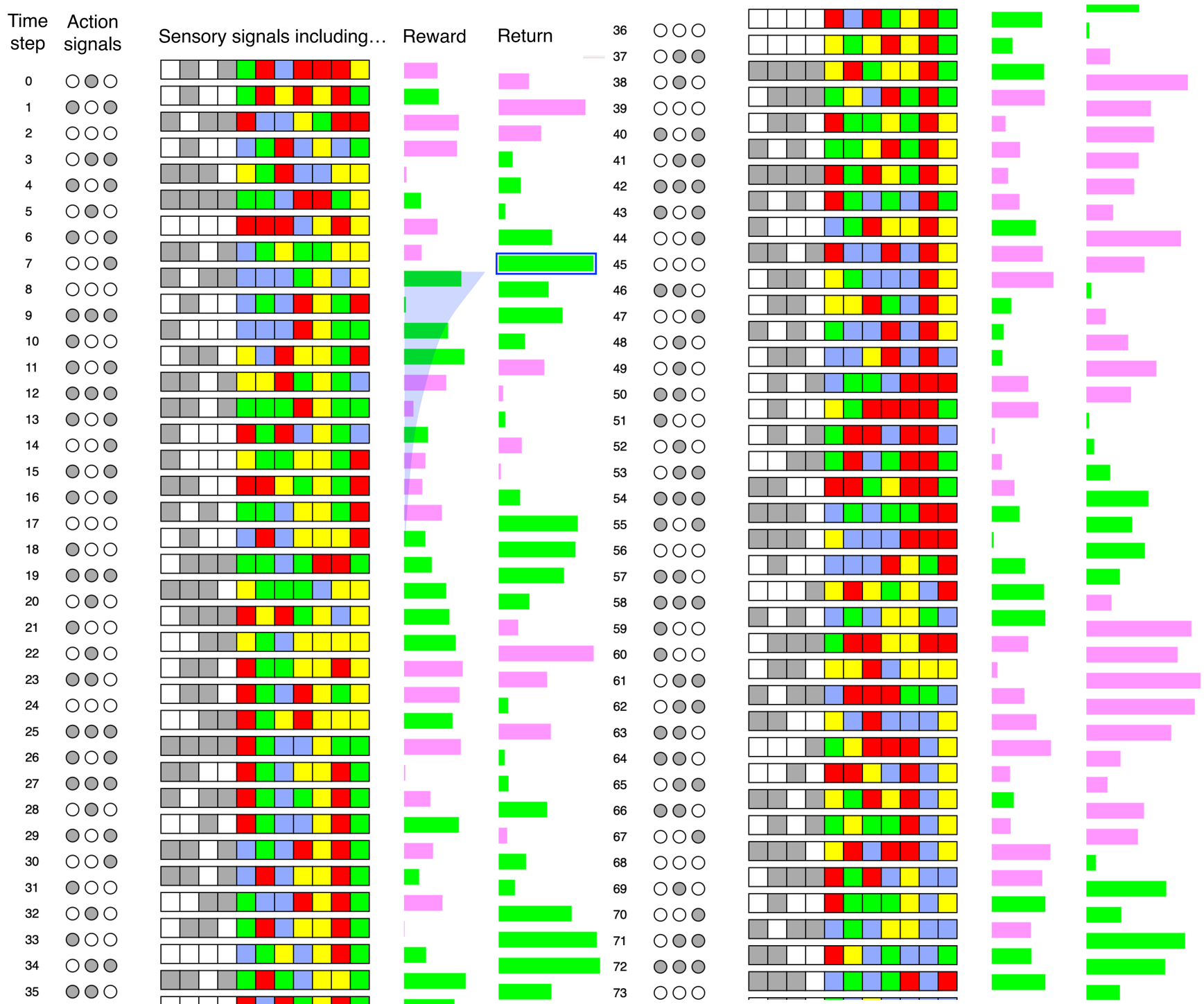
- Different sensory signals can be *qualitatively different* from each other
  - In their range of values
  - In their predictive relationships
    - to action signals
    - to each other
    - to themselves
- There are short-term *and* long-term patterns in these data
- There are many things to predict
- Prediction need not be just of the sensory signals
- The most important predictions are of *functions* of future sensory signals
  - e.g., predictions of *value*, the discounted sum of future reward
  - e.g., *General* value functions (GVFs)
    - predict any signal, not just reward
    - over a flexible temporal envelope
    - contingent on any policy
  - Predictions of different functions can *vary greatly* in their ability to be learned with computational efficiency



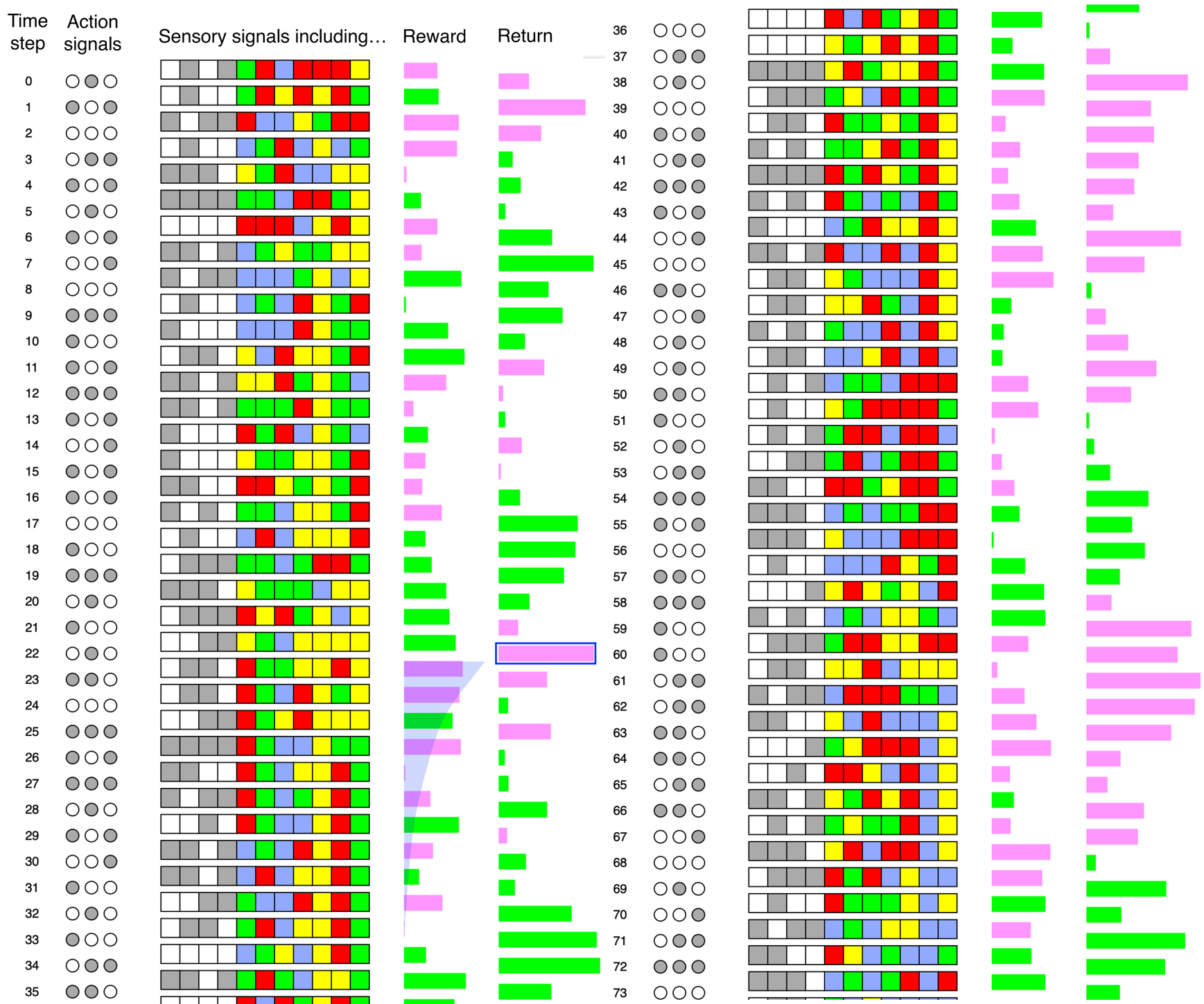
- Different sensory signals can be *qualitatively different* from each other
  - In their range of values
  - In their predictive relationships
    - to action signals
    - to each other
    - to themselves
- There are short-term *and* long-term patterns in these data
- There are many things to predict
- Prediction need not be just of the sensory signals
- The most important predictions are of *functions* of future sensory signals
  - e.g., predictions of *value*, the discounted sum of future reward
  - e.g., *General* value functions (GVFs)
    - predict any signal, not just reward
    - over a flexible temporal envelope
    - contingent on any policy
  - Predictions of different functions can *vary greatly* in their ability to be learned with computational efficiency



- Different sensory signals can be *qualitatively different* from each other
  - In their range of values
  - In their predictive relationships
    - to action signals
    - to each other
    - to themselves
- There are short-term *and* long-term patterns in these data
- There are many things to predict
- Prediction need not be just of the sensory signals
- The most important predictions are of *functions* of future sensory signals
  - e.g., predictions of *value*, the discounted sum of future reward
  - e.g., *General* value functions (GVFs)
    - predict any signal, not just reward
    - over a flexible temporal envelope
    - contingent on any policy
  - Predictions of different functions can *vary greatly* in their ability to be learned with computational efficiency



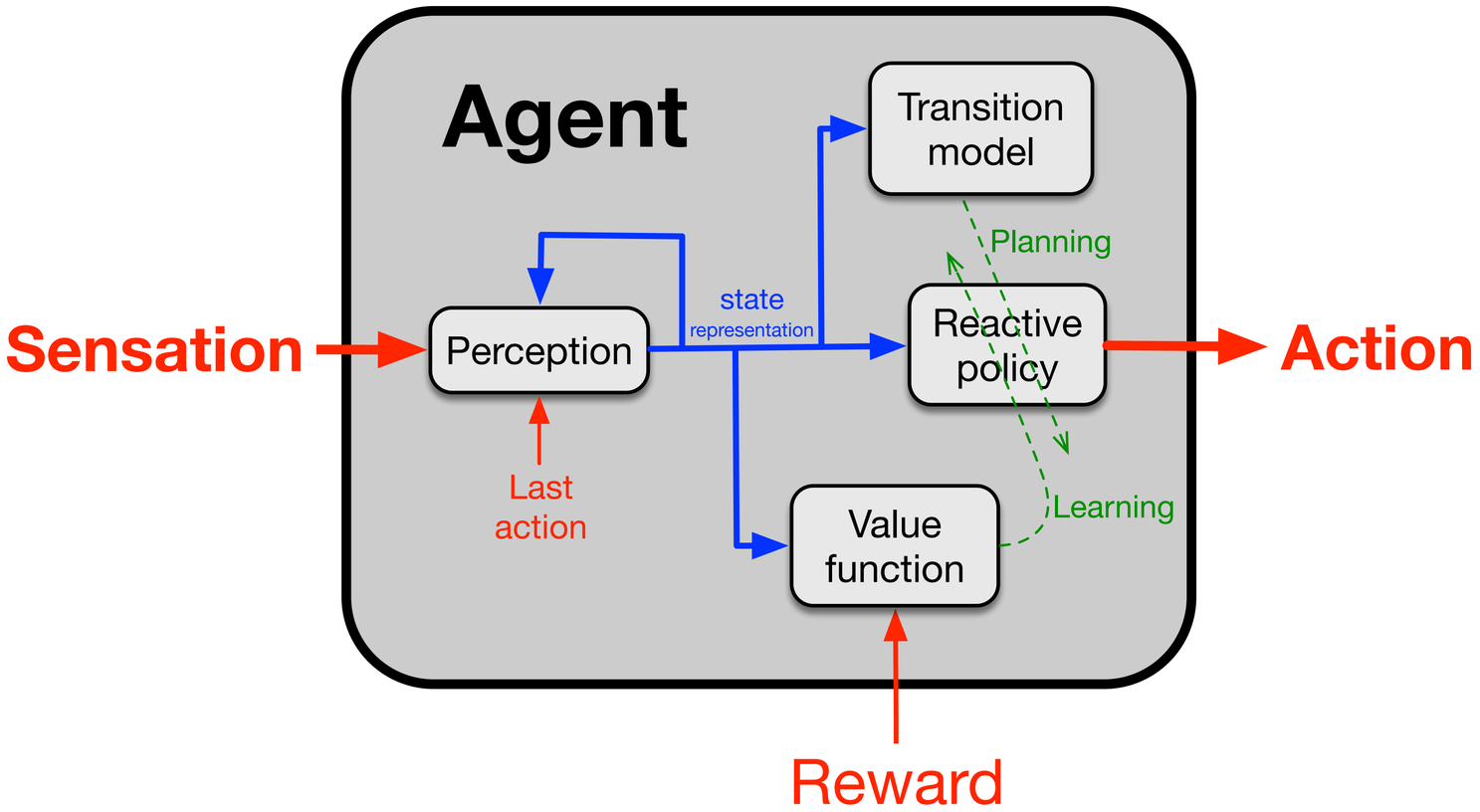
- Different sensory signals can be *qualitatively different* from each other
  - In their range of values
  - In their predictive relationships
    - to action signals
    - to each other
    - to themselves
- There are short-term *and* long-term patterns in these data
- There are many things to predict
- Prediction need not be just of the sensory signals
- The most important predictions are of *functions* of future sensory signals
  - e.g., predictions of *value*, the discounted sum of future reward
  - e.g., *General* value functions (GVFs)
    - predict any signal, not just reward
    - over a flexible temporal envelope
    - contingent on any policy
  - Predictions of different functions can *vary greatly* in their ability to be learned with computational efficiency



Now, let’s characterize the solution of intelligence
(i.e., take a peek inside the agent)

The common model of the intelligent agent

Common to RL, psychology, control theory, economics, neuroscience, operations research…



The agent comprises four components:

Perception produces the state representation used by all components

Reactive Policy quickly produces an action appropriate to the state

Value Function evaluates how well it's going, changes the policy (learning)

Transition model predicts the consequences of alternate actions, changes the policy (planning)

Distinctive features of the Alberta RL approach to AI

(how we keep our Eyes on the Prize)

1. We formulate intelligence as *real-time signal processing to maximize reward* (with limited computational resources)

2. Our agents learn continually from ordinary, unprepared experience (not from special training sets or demonstrations)

3. Our agents, though vastly complex, are much less complex than the world (the Big-World perspective) and thus must approximate, even in the limit

4. Our agents continually build abstractions, particularly time abstractions (options, GVFs) that enable them to think simply about large spans of time

“The Alberta Plan for AI Research”

A strategy document by Mike Bowling, Patrick Pilarski, and myself, coming soon to arXiv.
A 12-step plan to reach human-level AI ahead of schedule (before 2040)

1. Representation I: Continual supervised learning with given features
2. Representation II: Supervised feature finding
3. Prediction I: Continual GVF prediction learning
4. Control I: Continual actor-critic control
5. Prediction II: Average-reward GVF learning
6. Control II: Continuing control problems
7. Planning I: Planning with average reward
8. Prototype-AI I: One-step model-based RL with continual function approximation
9. Planning II: Search control and exploration
10. Prototype-AI II: The STOMP progression
11. Prototype-AI III: The Oak architecture
12. Prototype Intelligence amplification

Outline

- The Prize of understanding intelligence
- Why we seek it
- How reinforcement learning research in Alberta keeps its eyes on the Prize
    - by focusing on experience

“To make AI for good and for all”

Thank you for your attention