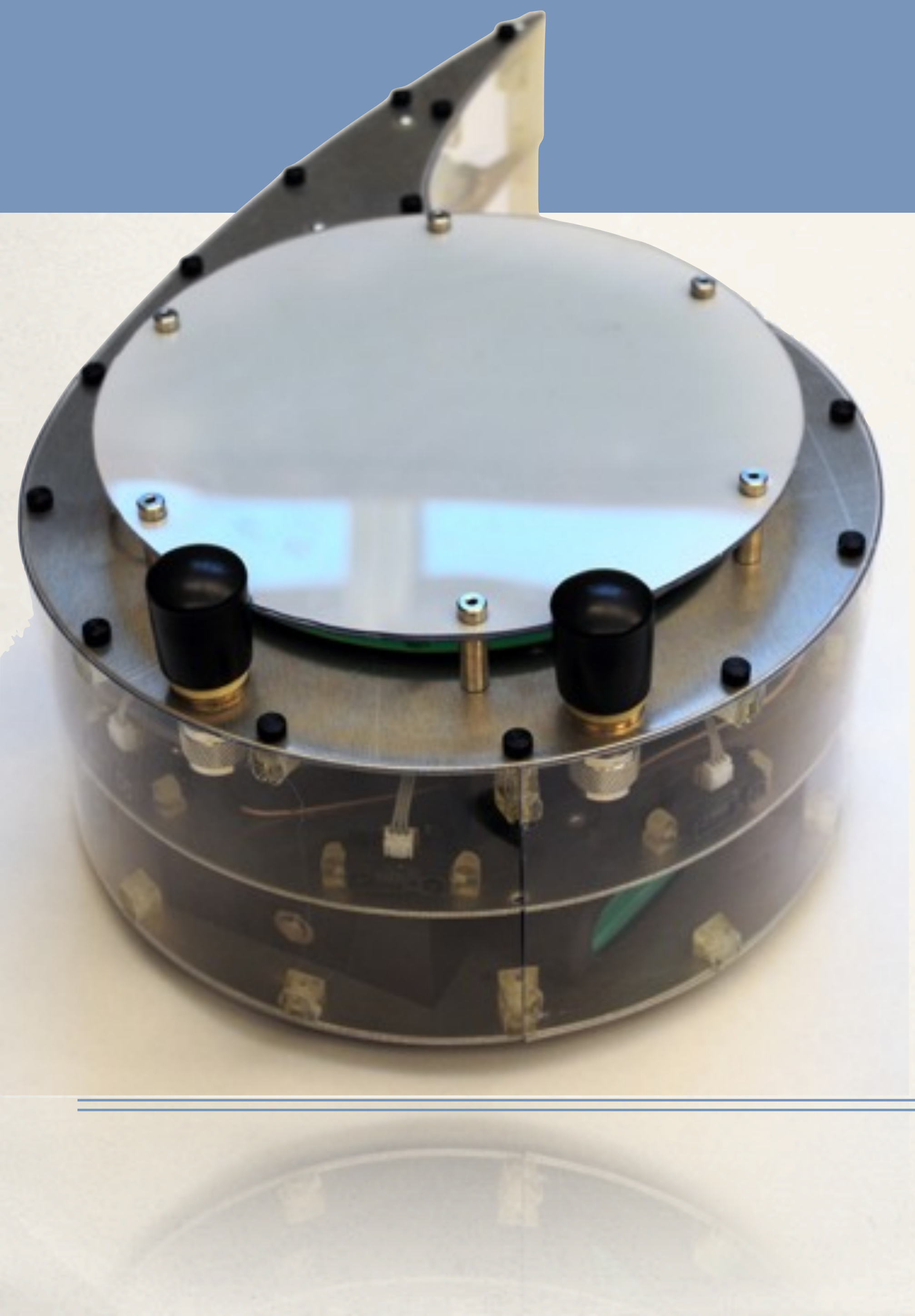
Reinforcement Learning Artificial Intelligence for Robotic Adaptation to New Environments



Patrick M. Pilarski
Postdoctoral Fellow, RLAI & AICML
University of Alberta

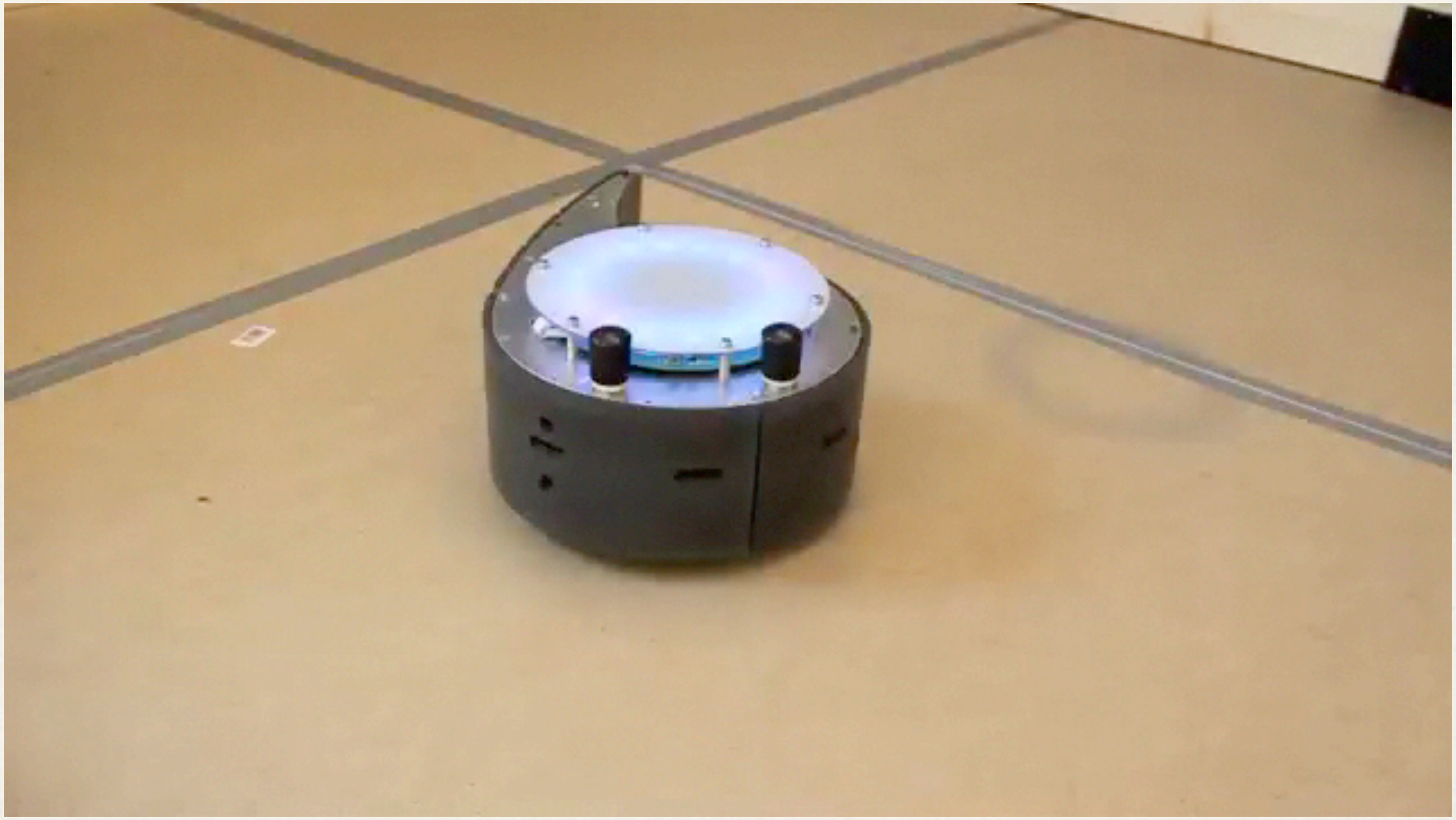
Joint work with Thomas Degris, Joseph Modayil, *and* Richard S. Sutton

November 3rd, 2010

Overview

- **The Critterbot Mobile Robotic Platform**
  - System overview.
  - Challenges for autonomy.
- **Reinforcement Learning Artificial Intelligence**
- **Applications of RL to a Mobile Autonomous Robot**
  - Prediction of stopping and impact times.
  - Control and adaptation in a changing environment.
- **Conclusions and Thoughts to Leave With**

The Critterbot: a Mobile Robot



http://critterbot.rl-community.org/home

Video by Thomas Degris

Key Challenges for Autonomy

- Extended operation times
- A Changing World
- A Changing Robot
- Changing Goals
- Prediction and Control

Key Challenges for Autonomy

- Extended operation times
- A Changing World
- A Changing Robot
- Changing Goals
- Prediction and Control

Open Questions for Research

1. How to best represent sensory information, and general knowledge about the world and the robot (and do this automatically).

2. How to detect and adapt to changes in the environment and robot.

3. How to make predictions and control policies for unexperienced environments in advance; i.e. learn about driving on fresh snow or ice while driving on wet pavement.

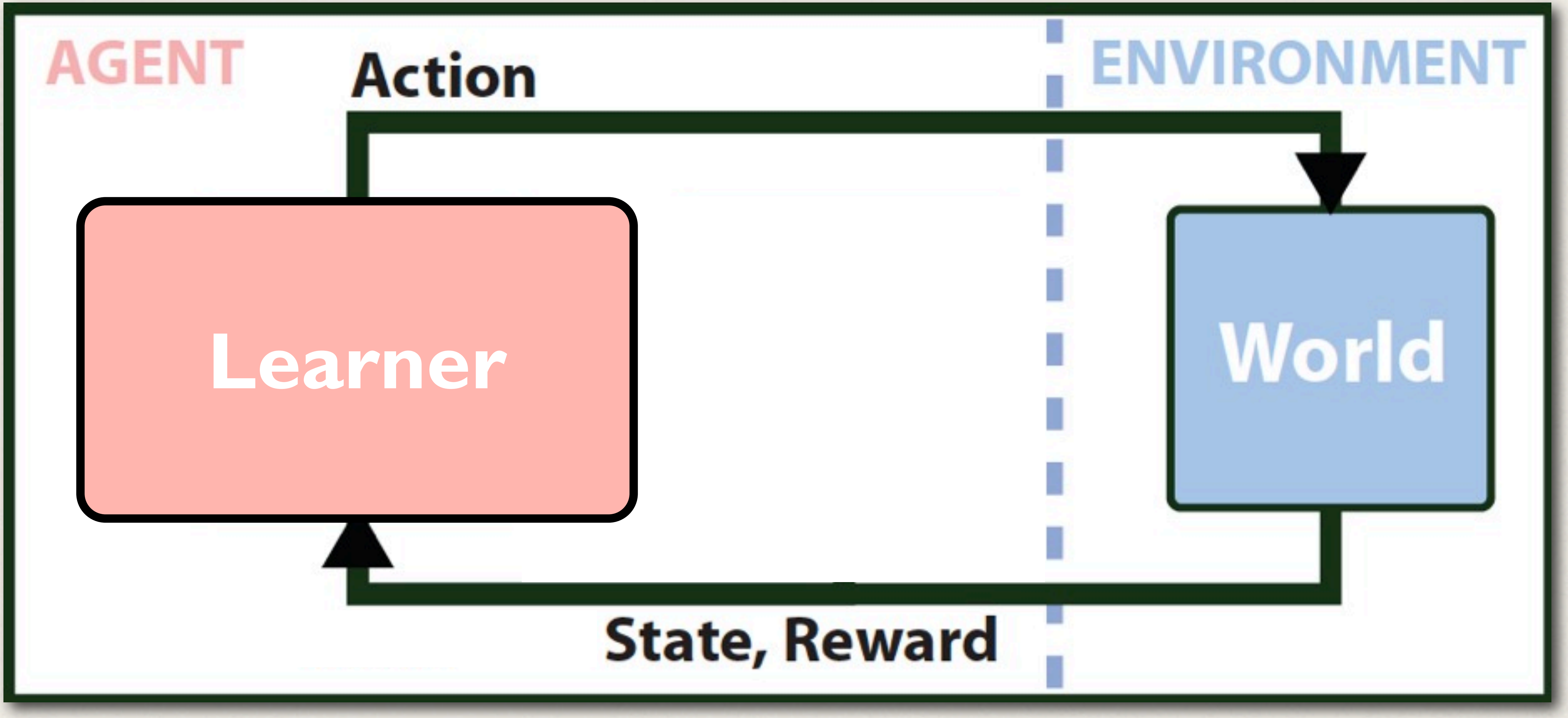
Machine learning allows us to directly address these questions.

Reinforcement Learning
Artificial Intelligence

- Reinforcement learning (RL) involves an **agent** and an **environment**.
- The agent perceives the state of the environment via a set of **observations** and takes **actions**.
- It then receives a new set of observations and a **reward** from the environment.
- These observations and rewards are used to predict future rewards, and change to the agent's **policy** (how it selects actions).
- **Key point:** RL methods involve **semi-supervised learning**. A single, scalar reward signal drives learning.

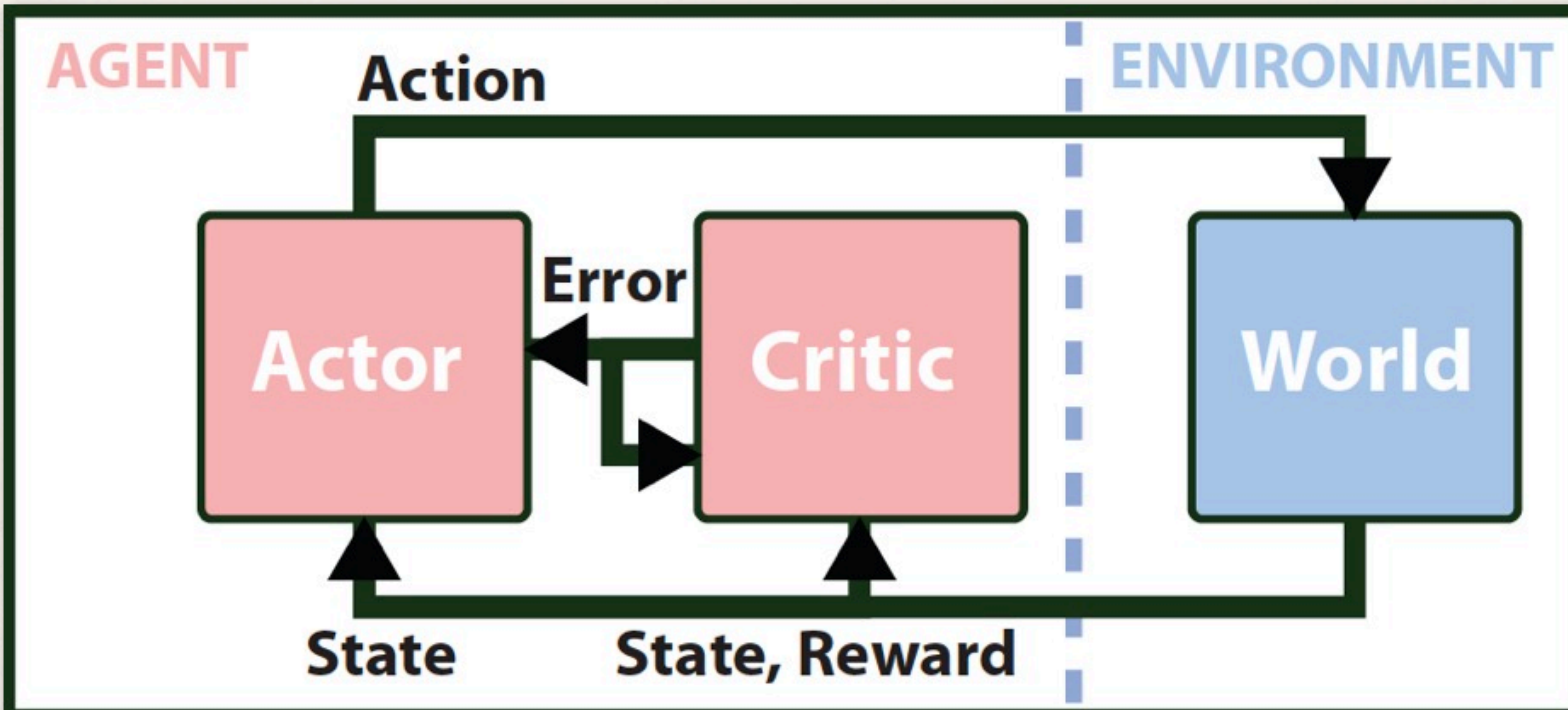
Sutton and Barto, MIT Press (1998)

Reinforcement Learning



Sutton and Barto, MIT Press (1998)

Reinforcement Learning



Sutton and Barto, MIT Press (1998)

What does this mean for real-world applications?

- RL systems can learn well when an end goal or desired behaviour is known but it is difficult (or impossible) to model the problem domain.
- Fast computation and low memory requirements allow for realtime deployment, especially on embedded or distributed systems.
- This also permits online adaptation: the learner can change in response to user needs and variation in the environment. This increases the robustness and versatility of systems.
- Very little hand tuning is required, and automatic tuning further reduces the need for ongoing maintenance. This saves human labour.

… and specifically for transport?

- These artificial intelligence techniques can be used in a general form *without* the requirement for detailed technical knowledge… objectives can be specified in simple terms (i.e. reward for fuel efficiency).

- **Online fault correction:** e.g., detecting and compensating for wear-and-tear, or a broken part.

- **Robustness to new conditions and environments:** e.g., quick adaptation to previously unseen weather or road conditions.

- **Safety:** rapid prediction of potentially dangerous events.

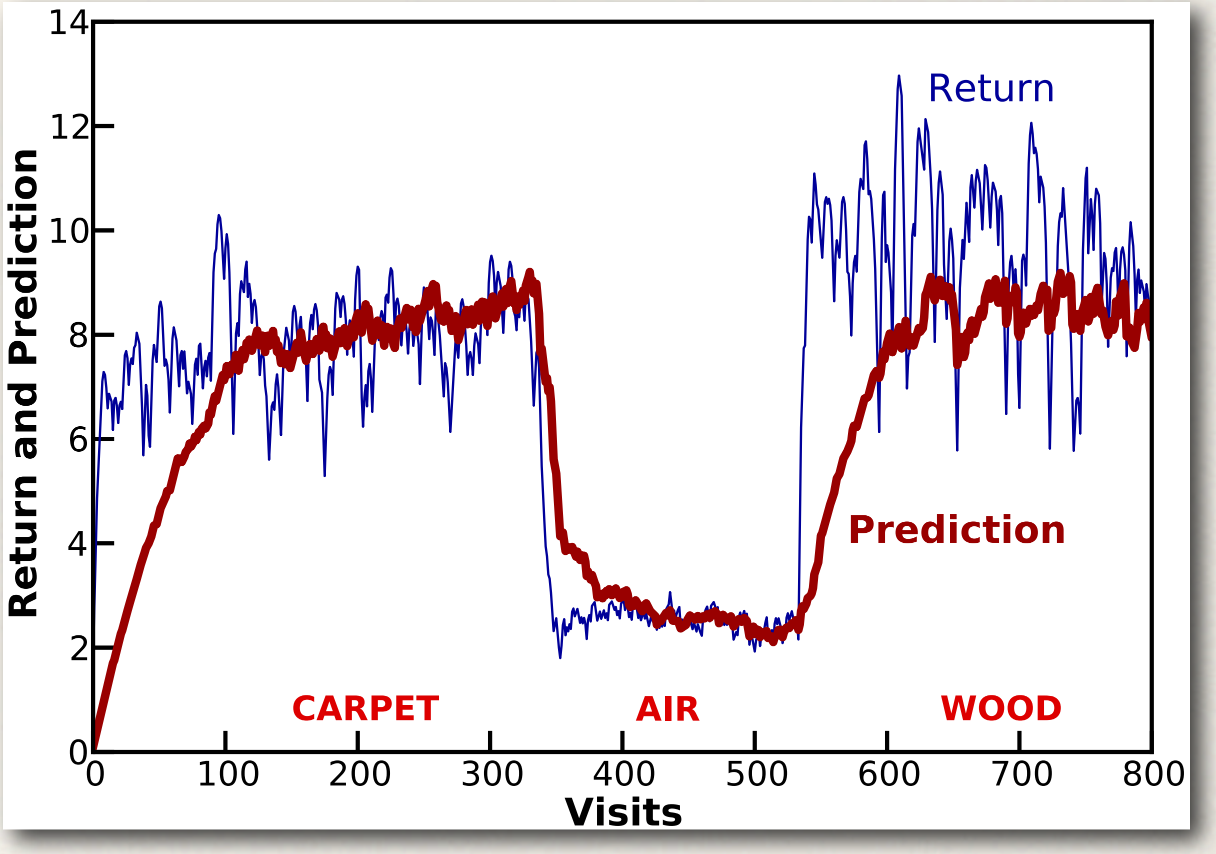
- Monitor motor **efficiency** and change behaviour to maximize it.

Two Examples: Movement

- The RL agent is able to observe the state of the robot and choose the actions (motor velocity commands).
- Through trial and error, it uses reward feedback to explore the space of possible behaviour policies.
- Learning is online and is done in real-time.

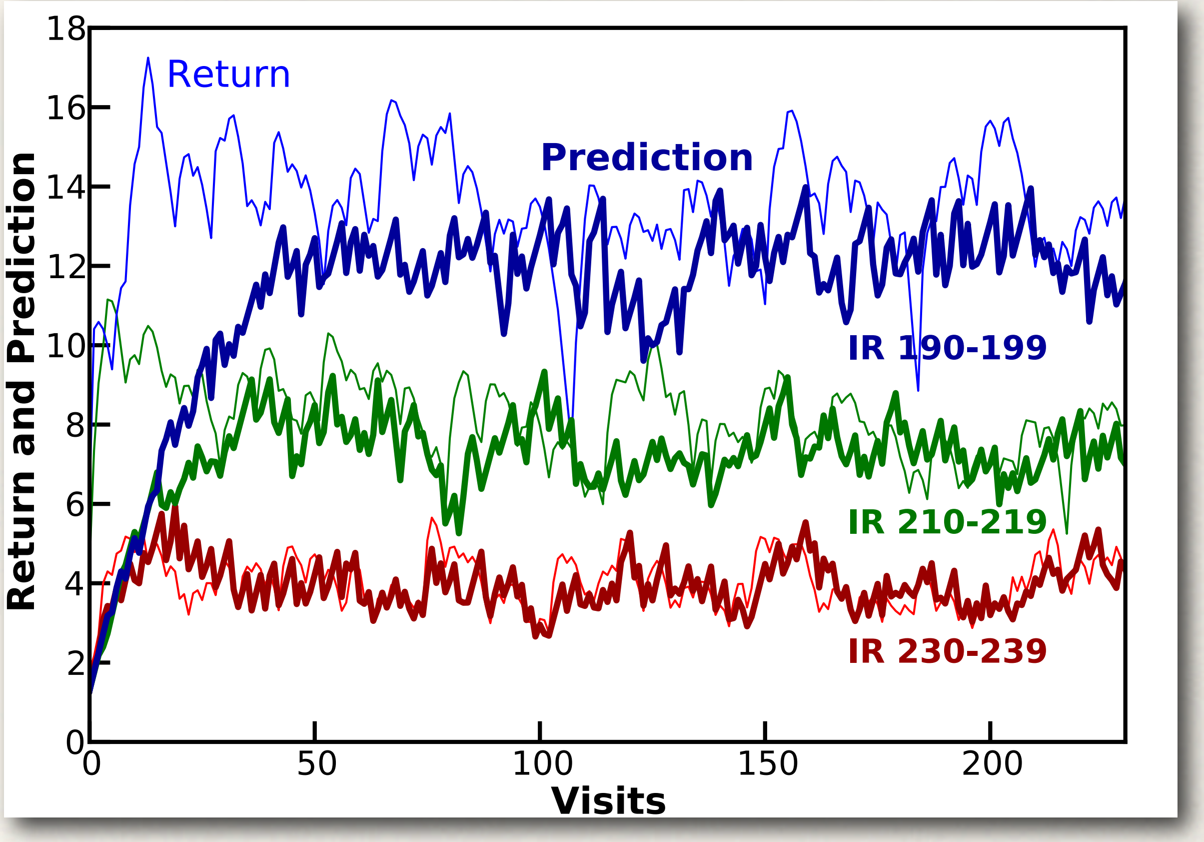
Prediction: Knowledge of Stopping and Impact Times

- **Representing knowledge in the form of general value functions.**
- **Key idea:** asking questions about the world.
- If I tried to stop right now, how long would it take?
- Given my current speed and sensors, how long until I hit the wall?



Prediction: Knowledge of Stopping and Impact Times

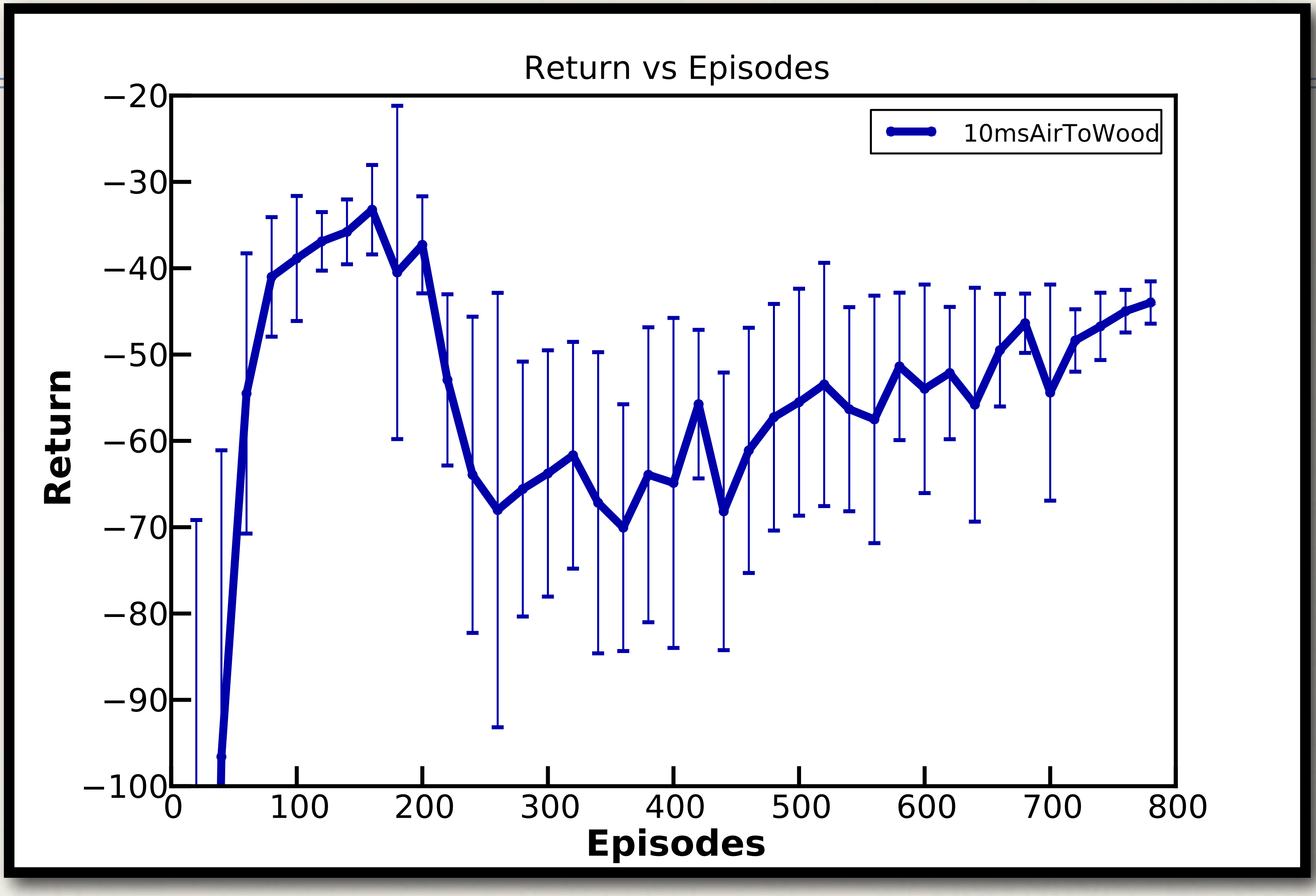
- **Representing knowledge in the form of general value functions.**
- **Key idea:** asking questions about the world.
- If I tried to stop right now, how long would it take?
- Given my current speed and sensors, how long until I hit the wall?



Control: Adapting to New Environments

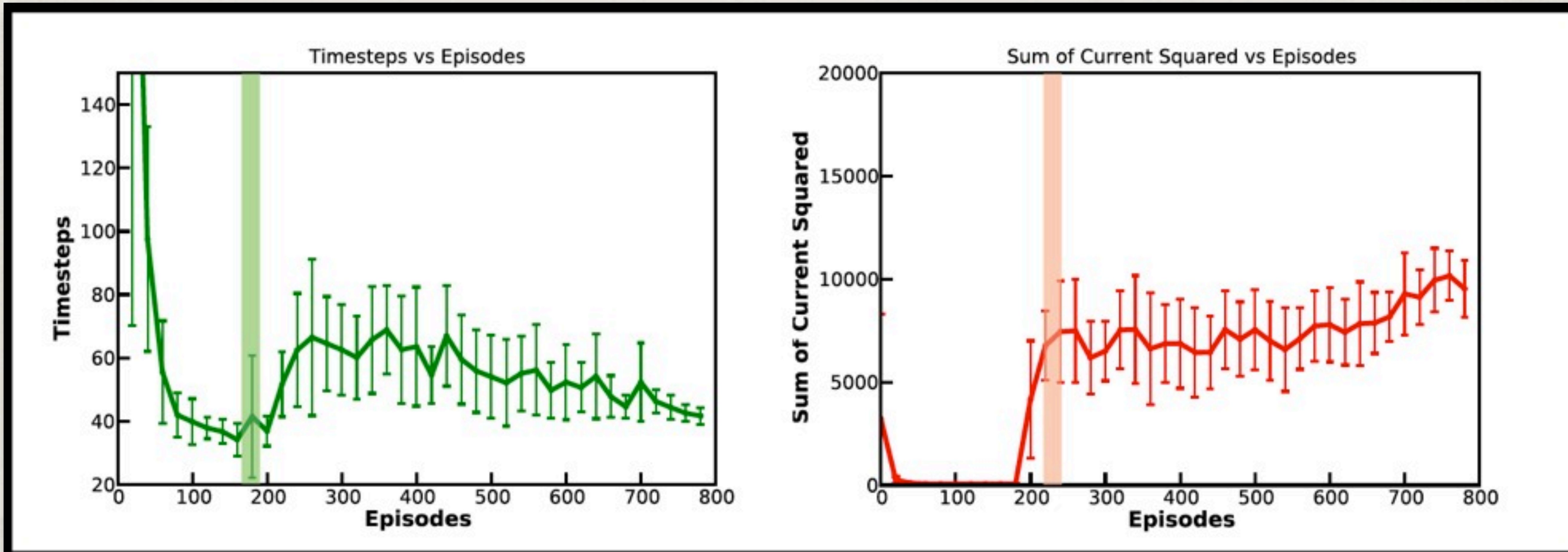
- Task of balancing the current (power) drawn by its motors with the aim of reaching a desired wheel velocity in as short a time as possible.
- The agent receives a negative reward on every timestep, and is also penalized in proportion to its wheel current measurements.
- Learns for 800 episodes. After Episode 200, robot transitioned from suspended (in air) to grounded (on a wooden surface) motion.
- On the ground, the robot's mass and wheel friction played additional roles in determining the current usage, making the trade-off between power and acceleration length more complex.

Control Results



- The agent was able to successfully adapt to the environmental change. Continuous online adaptation allows behavioural change following unexpected events or system perturbations.

Control Results



- The agent was able to successfully adapt to the environmental change. Continuous online adaptation allows behavioural change following unexpected events or system perturbations.

Key Messages to Leave With

- Reinforcement learning artificial intelligence methods are well suited to problems where the environment may be variable, complex, or poorly defined (semi-supervised & flexible online learning).
- Adaptive control methods of this type allow a mobile robot to build up knowledge about itself / the world and use this knowledge to act.
- This enables new approaches to fault tolerance, power and fuel efficiency, and safety.
- While this is demonstrated on a small robotic system, these methods are expected to transfer well to large-scale mobile applications.

Acknowledgements

- **Collaborators:**

  Thomas Degris (*CS, U of Alberta*)
  Joseph Modayil (*CS, U of Alberta*)
  Richard S. Sutton (*CS, U of Alberta*)
  Marc G. Bellemare (*CS, U of Alberta*)

- **Funding and Support:**

  Alberta Ingenuity Centre for Machine Learning
  Alberta Innovates – Technology Futures (iCORE)
  Natural Sciences and Engineering Research Council of Canada