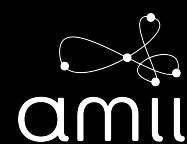


# UNIFYING PERSPECTIVES ON INTELLIGENCE: WHAT RL ADDS TO THE COMMON MODEL OF THE AGENT

**Scolol & ISAB Summer School Spotlight Talk**  
**23 August 2023**

Abhishek Naik



UNIVERSITY OF  
ALBERTA



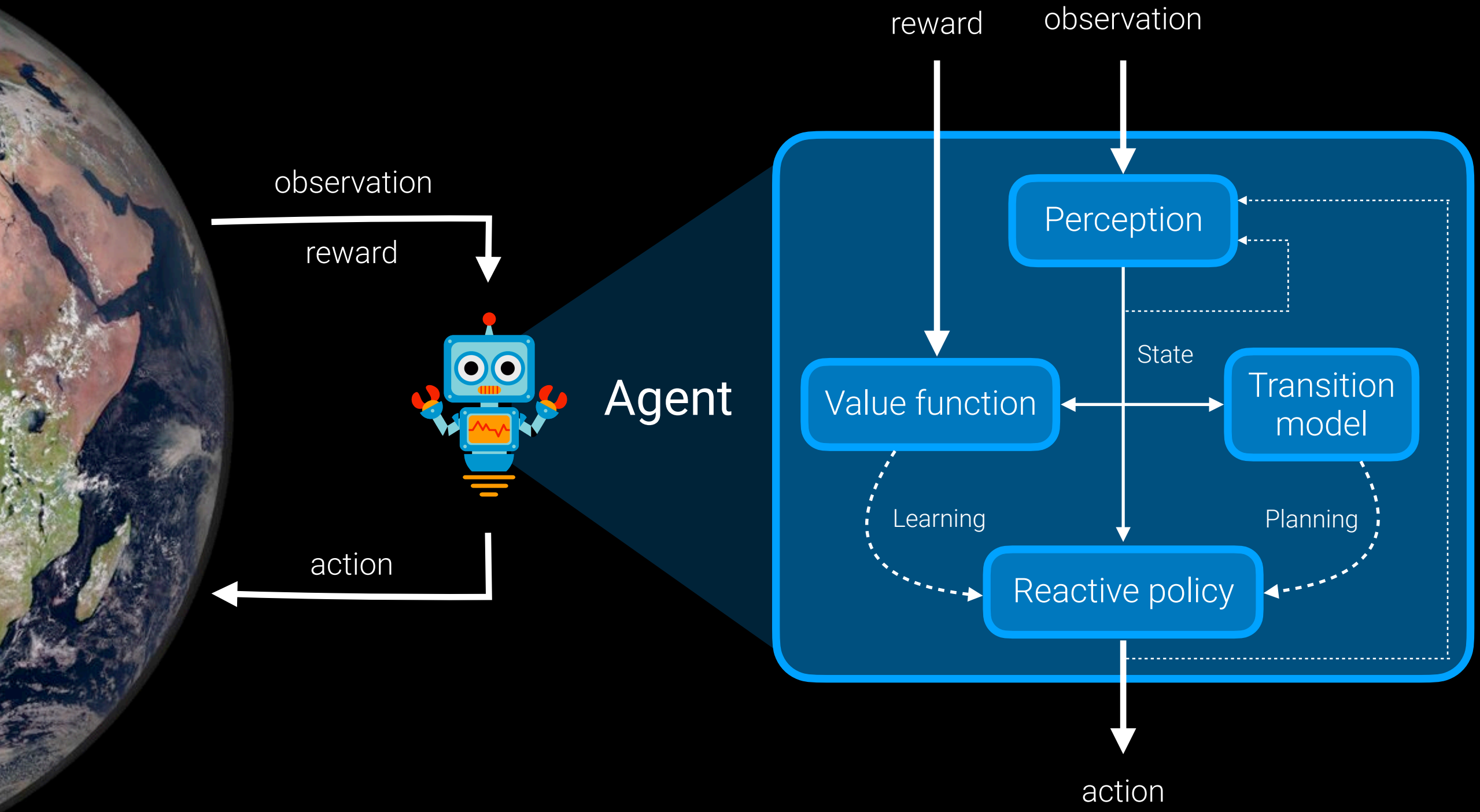
# A WORKING DEFINITION OF INTELLIGENCE

“Intelligence is the computational part of  
the ability to achieve goals”

- John McCarthy

non-computational parts: being stronger, having better sensors

# THE COMMON MODEL OF THE DECISION-MAKER



# THE REWARD HYPOTHESIS

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

- Michael Littman, Rich Sutton

# TEMPORAL-DIFFERENCE LEARNING: AN ALGORITHM TO MAXIMIZE LONG-TERM REWARD

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{correct})$$

$$P_{new} = (1 - \alpha)P_{old} + \alpha(P_{better})$$

$$= P_{old} + \alpha(P_{better} - P_{old})$$

$$V_{new}(s) = V_{old}(s) + \alpha(\underbrace{R + V_{old}(s') - V_{old}(s)}_{\text{TD error}})$$

TD error

inspired from psychology and constrained by computation

# TD LEARNING BEST FITS VARIOUS PSYCH/NEURO DATA

- ▶ explains blocking and higher-order conditioning
- ▶ predicted the reversal of blocking — later confirmed by Kehoe et al. (1987)
- ▶ experimental support for the reward-prediction-error hypothesis: Schultz et al. (1997)
- ▶ causal support using optogenetics: Steinberg et al. (2013)

# TAKEAWAYS

- ▶ There are many commonalities in how our related fields are thinking about the phenomenon of intelligent behavior.

Let's use some common terminology to foster more collaborations!

- ▶ To this common model of the agent, RL adds:
  - ▶ the reward hypothesis,
  - ▶ general and scalable algorithms to maximize reward, some of which are biologically plausible.



My poster is about a more efficient variant of TD-learning, specific to lifelong learning!  
(Group A)

# REFERENCES

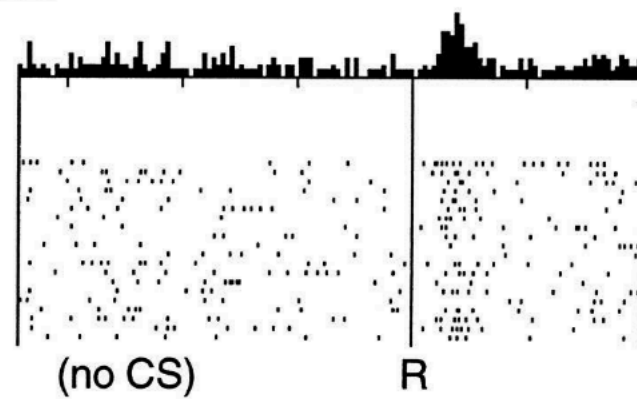
- ▶ Kehoe, E. J., Schreurs, B. G., & Graham, P. (1987). Temporal primacy overrides prior training in serial compound conditioning of the rabbit's nictitating membrane response. *Animal Learning & Behavior*.
- ▶ Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*.
- ▶ Ludvig, E. A., Bellemare, M. G., & Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications*.
- ▶ Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*.
- ▶ Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. *MIT Press*.
- ▶ Sutton, R. S. (2022). The Quest for a Common Model of the Intelligent Decision Maker. *Multi-disciplinary Conference on Reinforcement Learning and Decision Making (RLDM)*.



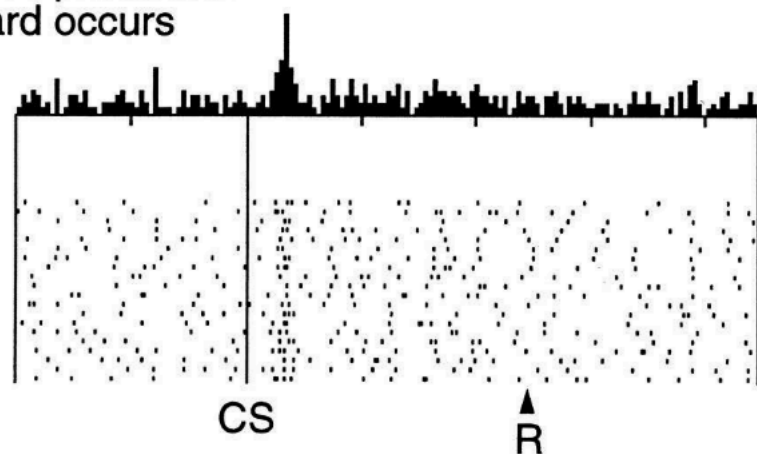
# THANK YOU

 [abhisheknaik96.github.io](https://abhisheknaik96.github.io)  
 [abhishek.naik@ualberta.ca](mailto:abhishek.naik@ualberta.ca)  
 [anaik96](https://twitter.com/anaik96)

No prediction  
Reward occurs



Reward predicted  
Reward occurs



Reward predicted  
No reward occurs

