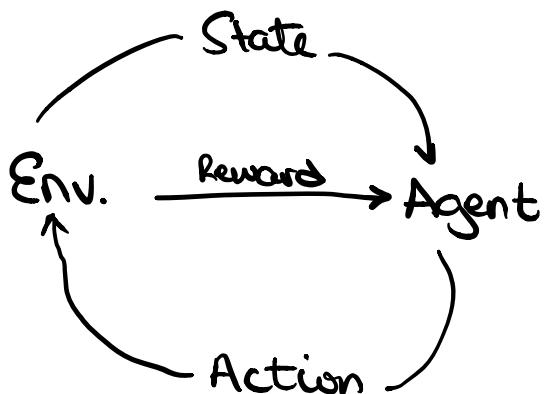


- Mindbridge is NOT a moonlanding company LOL.  
↳ emphasis on not.
- Mindbridge helped Ottawa 6<sup>th</sup> graders to create AI models for Lunar landing spacecrafts. Models were put onto microcontrollers with limited memory.
- The controllers will be put into space brought back to Earth to check for cosmic damage (from rays or radiation).
- The AI models were trained with Reinforcement Learning (RL)

Popular Books on RL:

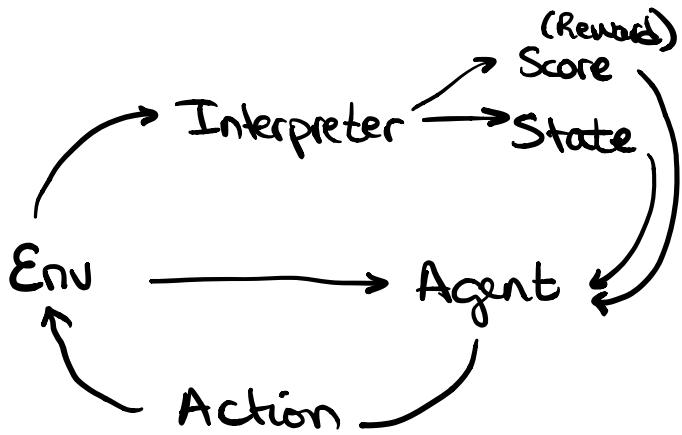
1. Reinforcement Learning — Sutton (De facto Standard)
2. Algorithms for Reinforcement Learning — Szepesvari (Math focused)

### RL AGENT - ENVIRONMENT RELATIONSHIP



>To get a model you are looking for RL@home will on the lowerbound require a GPU 6-24 hours.

## RL INTERPRETER-AGENT-STATE



Mindbridge built AI model to play street fighter.

⇒ wanted to be superhuman.

humans react 50-100ms response time.

Used,

- Darknet YOLO for human detection

- OpenCV for health bar

⇒ • 30\$ webcam

Model worked but not

better than avg. human.

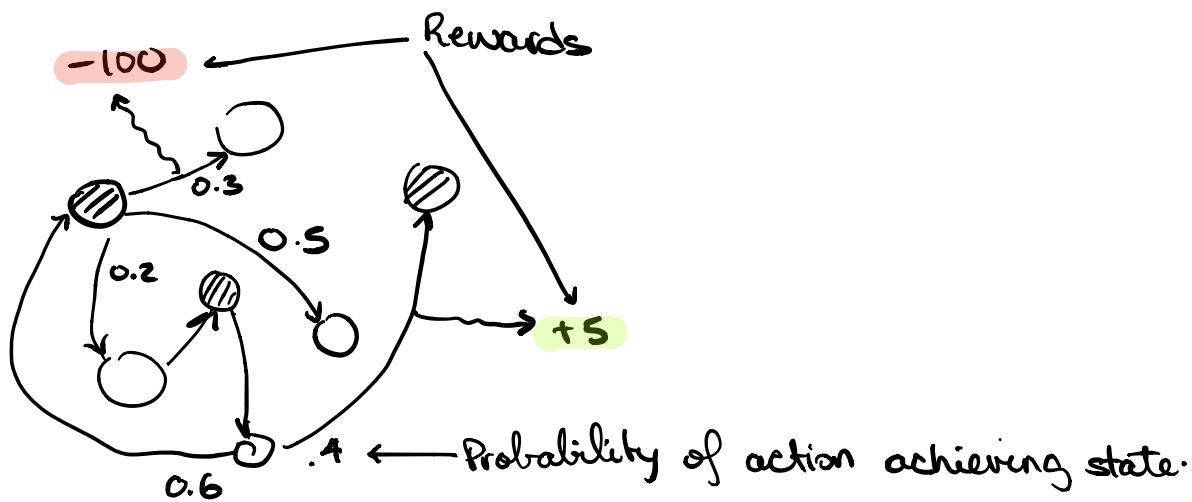
Described this RL for (2) reasons:

1. what is RL
2. difficulties w- employing RL.

## Markov Decision Processes (MDP)

Reinforcement learning is kind of like MDP

- the system (+ environment) have randomness.
- actions are connected to states probabilistically
- maximum reward often not immediate



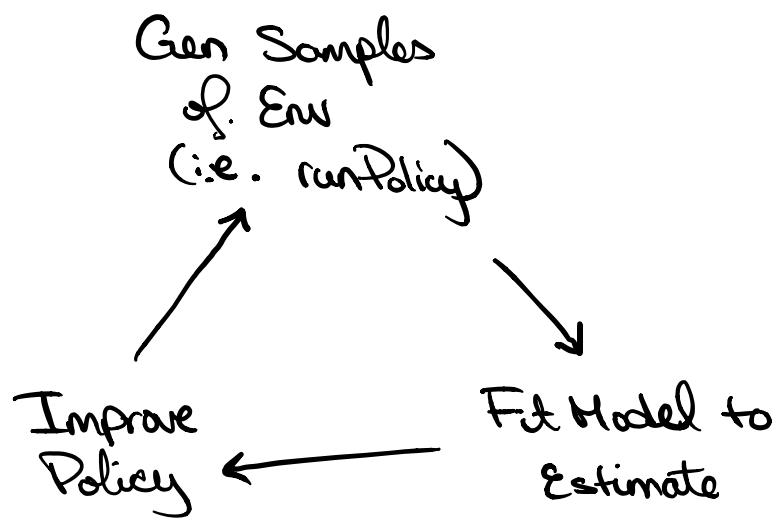
Policy Gradient is best for Control Networks.

- slow to converge
- high variance

In theory PG beats DQN.

DQN is deep Q learner.

- Q Table [state, action]
- Explore Stage (rand actions)
- Q Table Update Learning Table.



DQN uses future rewards to influence actions.