Mario, State Abstractions and the Wonderful World of Options

Nick Abou Risk and Richard Gibson

CMPUT 609 Project
University of Alberta

April 7, 2009

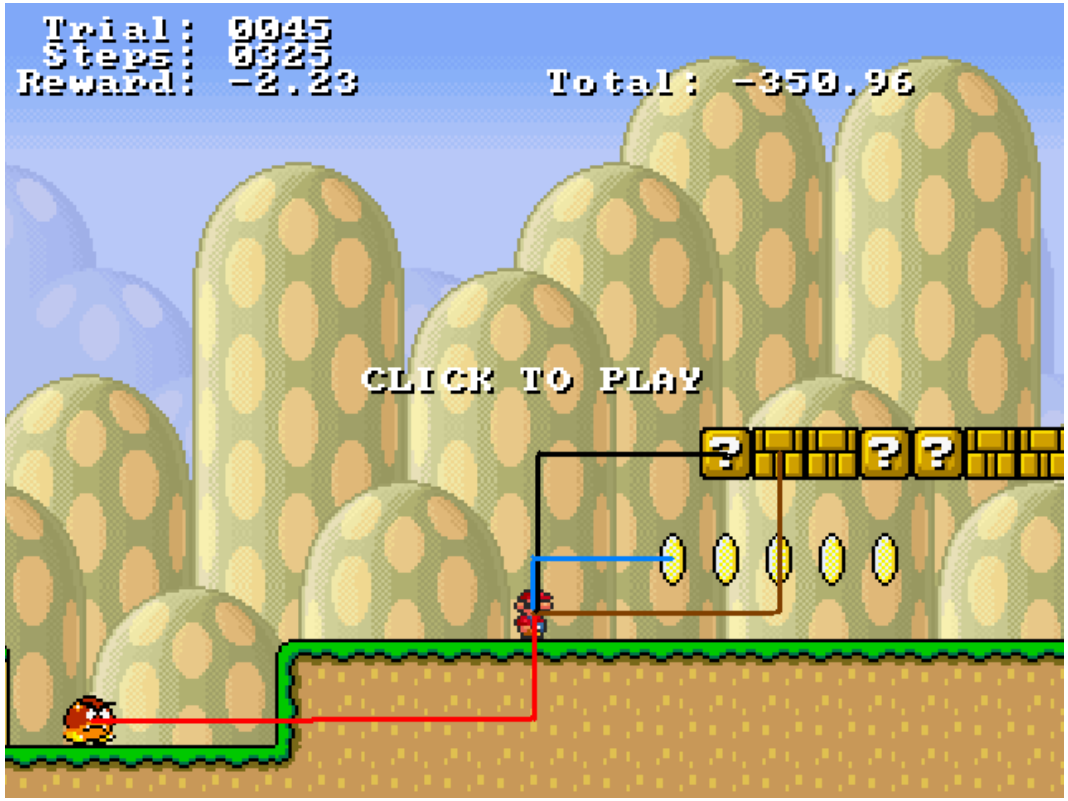
State Abstraction

Observation (State)

- Character array
- Position and velocity of each Monster
- Type and wingedness of each Monster

State Abstraction

- Full Observation-space is too large
- Features extracted from the Observation
- Most of these features are in egocentric representation
- Other features include Mario's size, number of enemies, etc.



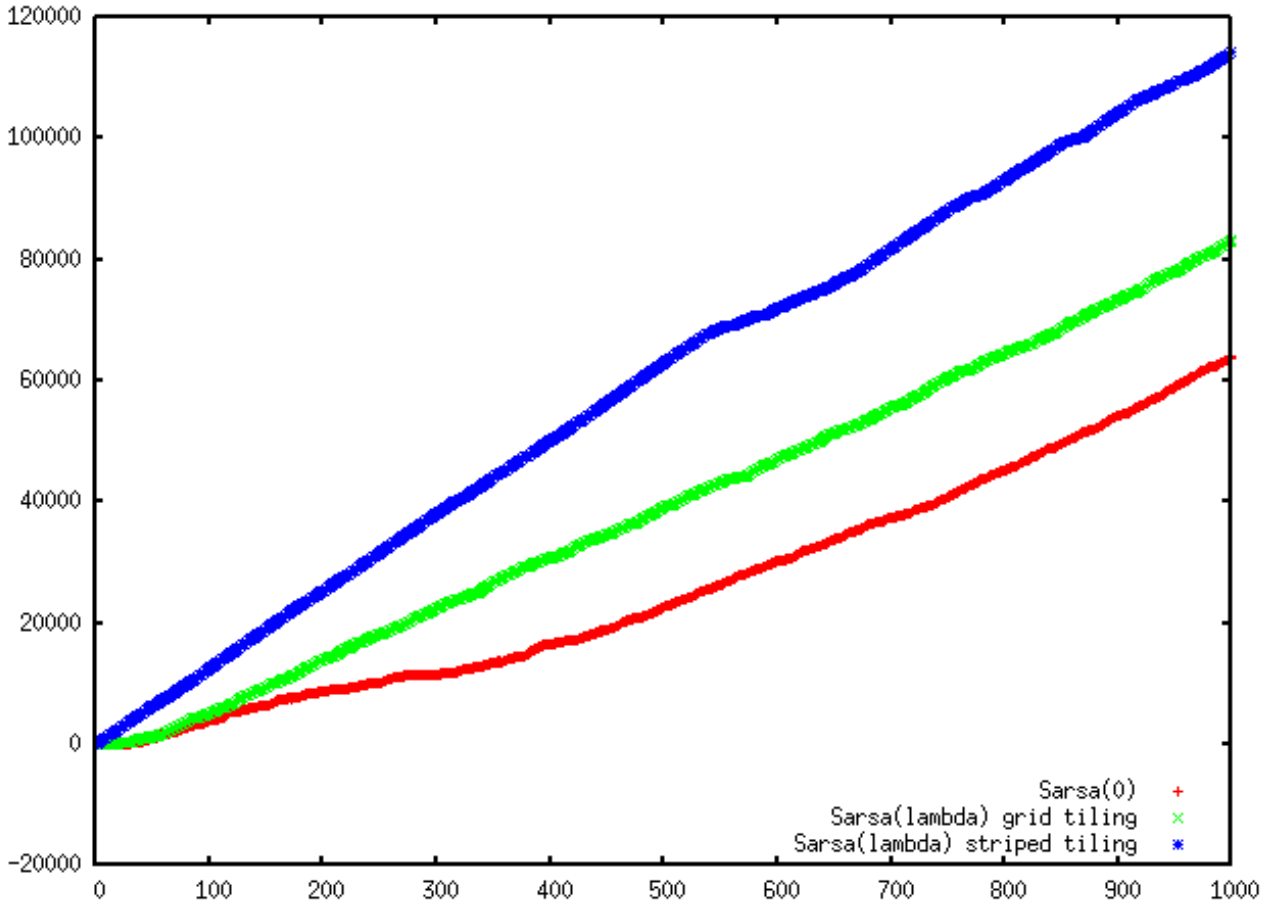
- small : Mario's size
- right : Mario's x-direction
- (-7,-3) : (x,y)-distance to nearest enemy
- (3, 1) : (x,y)-distance to nearest coin
- (4, 3) : (x,y)-distance to nearest question block
- (5, 3) : (x,y)-distance to nearest brick

Learning Algorithms

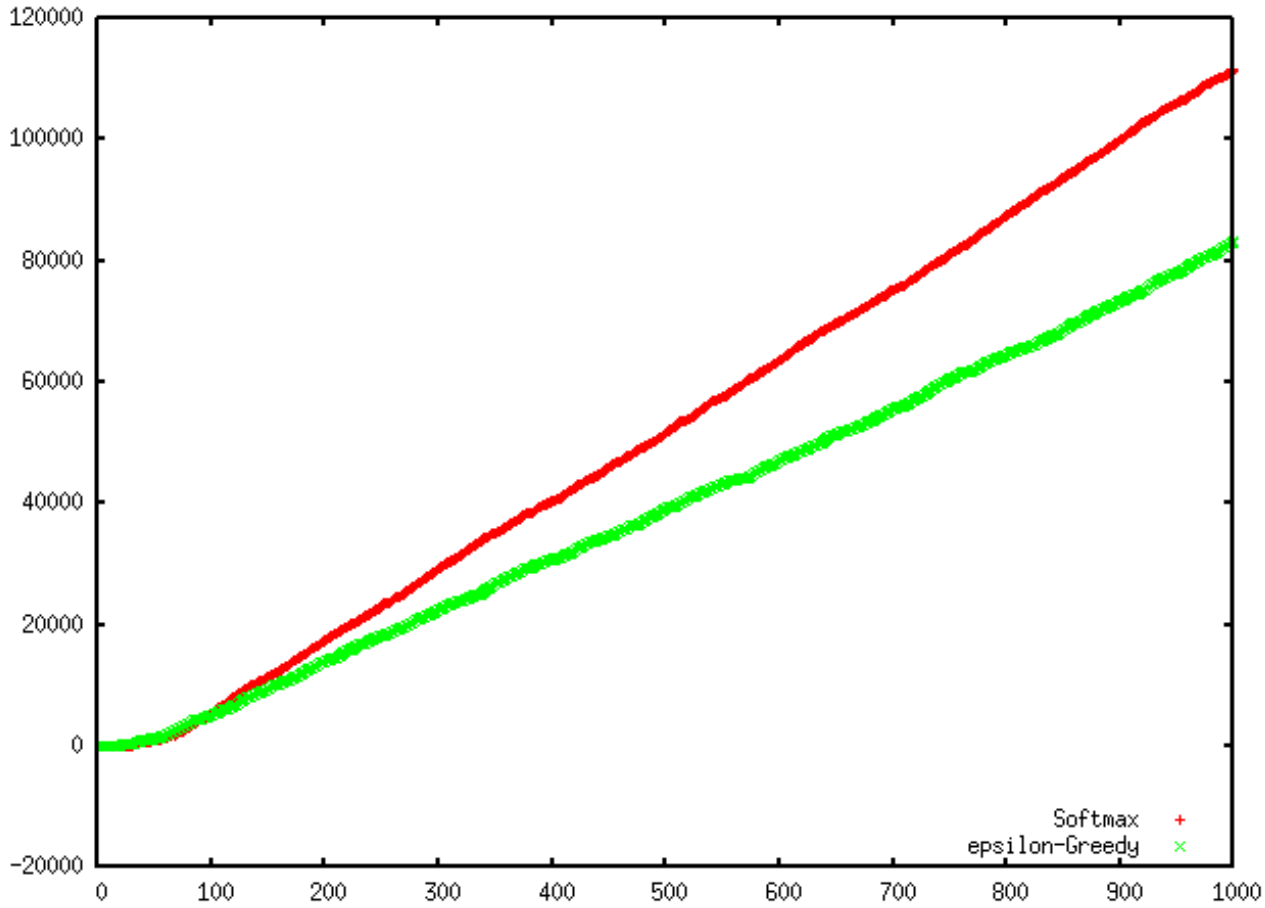
- Sarsa(0)
- Sarsa(λ)
- Watkin's Q-learning
- Option Learning

Policy Methods

- ϵ-Greedy
- Softmax



- Total reward earned by Sarsa(0), Sarsa(λ) with grid tiling, and Sarsa(λ) with striped tiling .
- Features include Mario's direction and distances to the finish line and nearest coin, question block, and enemy.

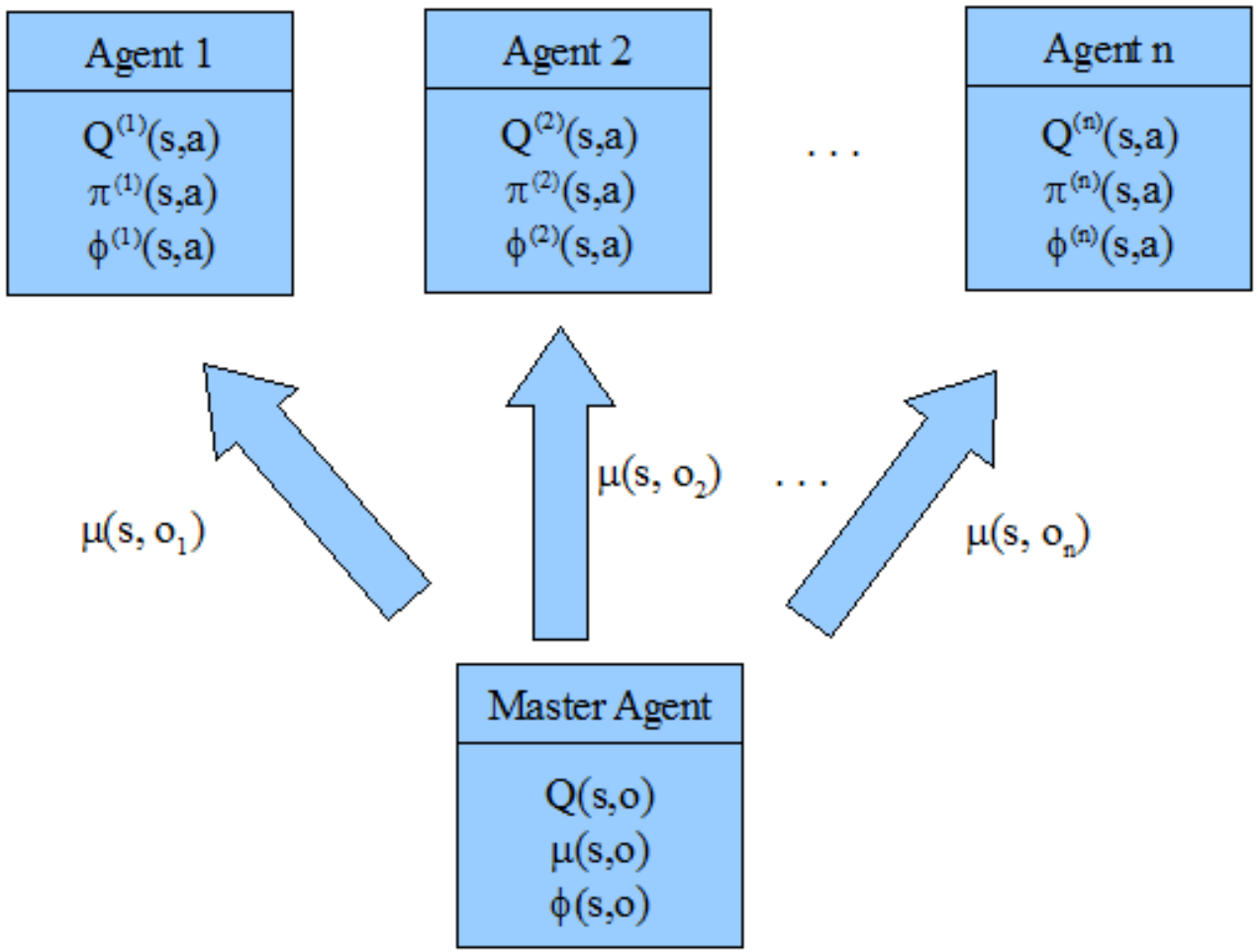


- Total reward earned by Softmax and ϵ-Greedy policies.
- Used Sarsa(λ) with grid tiling.
- Features include Mario's direction and distances to the finish line and nearest coin, question block, and enemy.

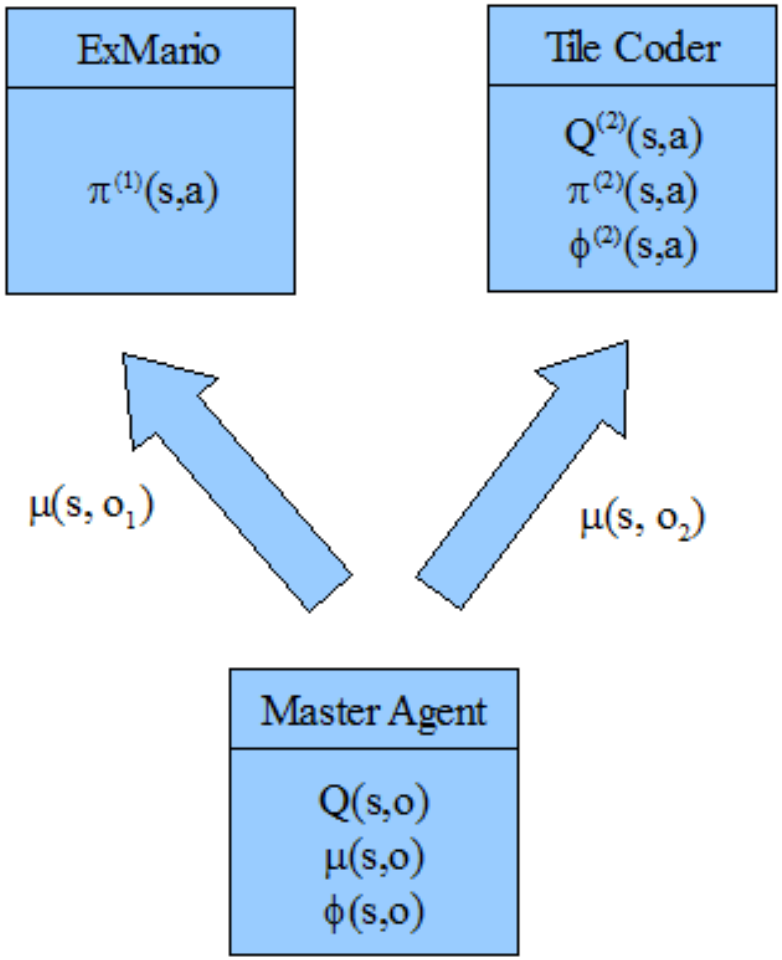
Intermission (video)!

- Trained for 1000 episodes
- Sarsa(λ) with striped tiling
- ϵ-Greedy policy
- Features include:
  - Mario's direction
  - Mario's size
  - Number of enemies on-screen
  - Nearest enemy's direction, type, and wingedness
  - Distance to finish line
  - Distance to nearest coin
  - Distance to nearest question block

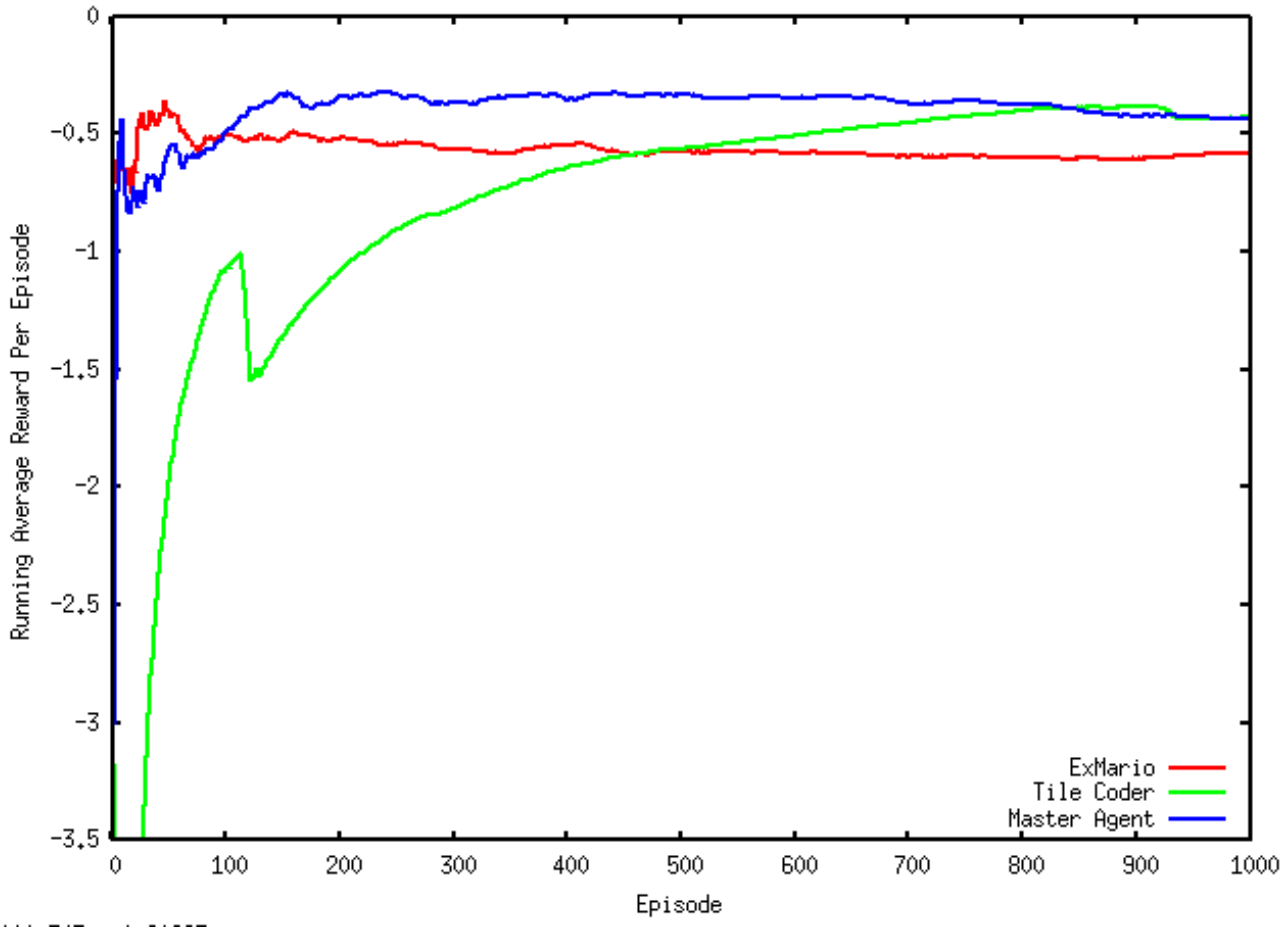
Options



- Each agent receives $(s_t, a_t, r_{t+1}, s_{t+1})$ at time step $t+1$ and learns off policy.

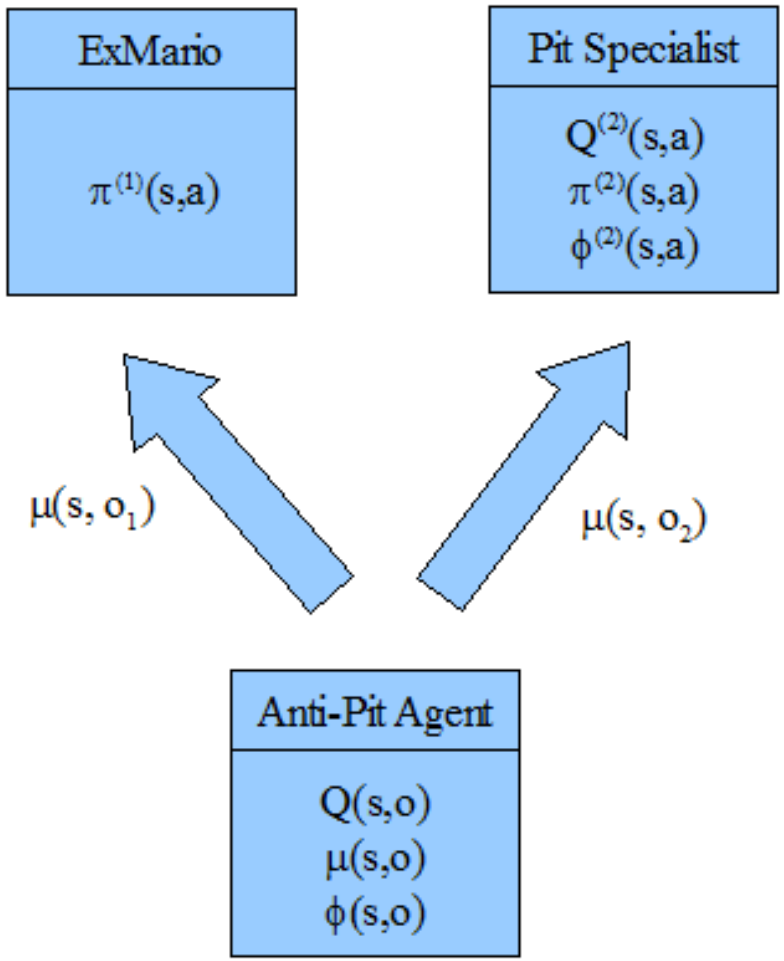


- Each of the 2 options available in every state.

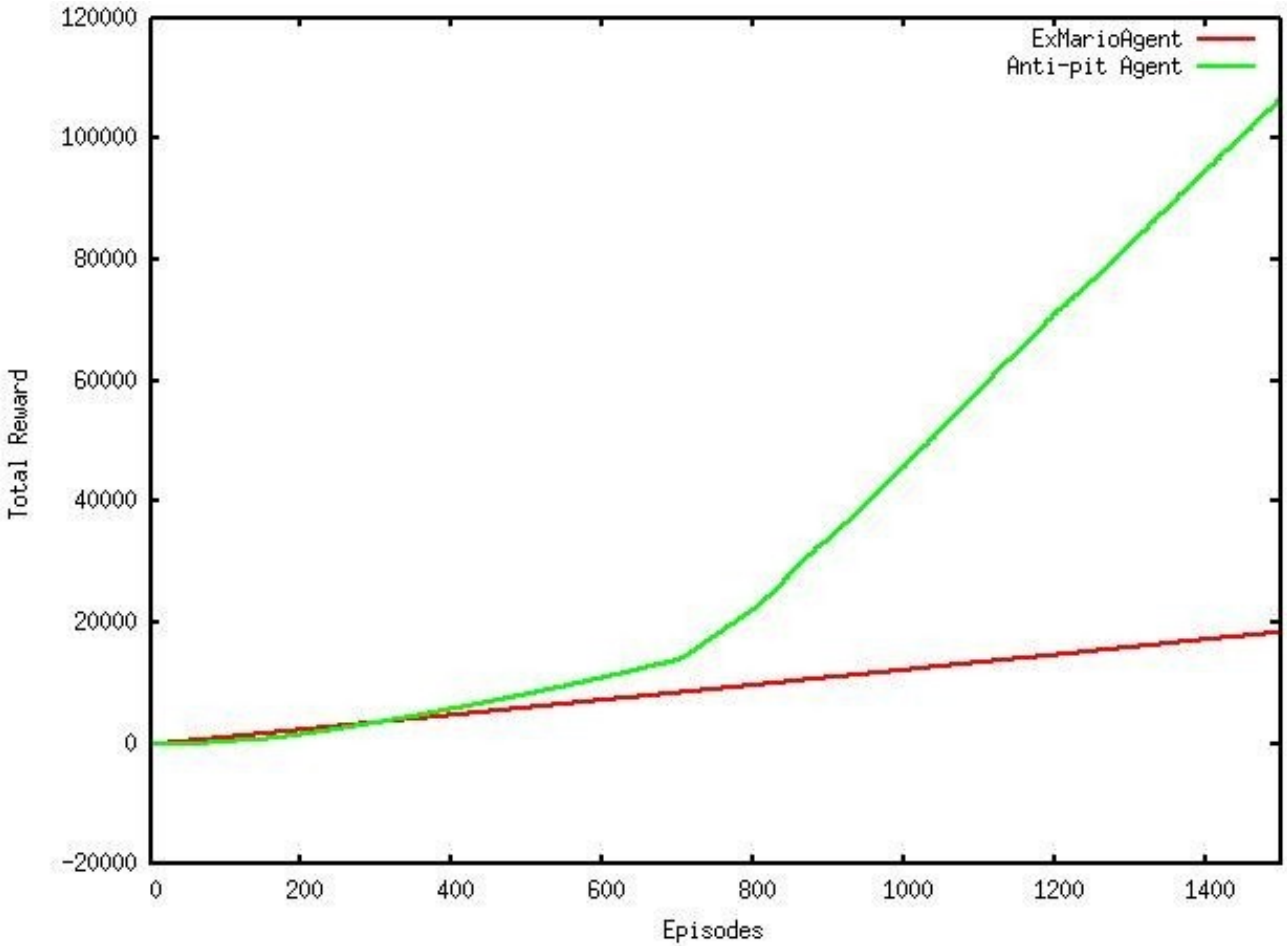


- Running average reward per episode for each ExMarioAgent (alone), our Tile Coder agent (alone) and the Master Agent which uses both as options.
- Run over 1000 episodes on a level where positive reward is difficult to achieve (just receive +1 reward for killing an enemy).

ExMarioAgent on difficulty 8 of 10.



- Pit Specialist option only available in states with a pit near Mario.



- Total reward earned by ExMarioAgent and Anti-Pit Agent on difficulty 8 of 10.
- Averaged over 2 runs of 1500 episodes each.

Anti-Pit Agent video

Things we learned...

Things we learned...

- Creating good state abstractions can be very time consuming and often involve a lot of tweaking.

Things we learned...

- Creating good state abstractions can be very time consuming and often involve a lot of tweaking.
- Options can be used as a powerful way to combine different state abstractions with differing strengths and weaknesses.

Things we learned...

- Creating good state abstractions can be very time consuming and often involve a lot of tweaking.
- Options can be used as a powerful way to combine different state abstractions with differing strengths and weaknesses.
- Mario is fun!

Thanks for listening!