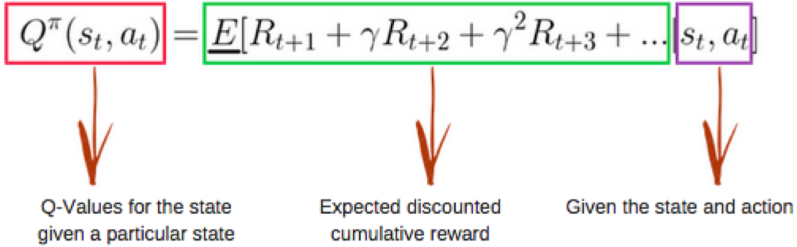
Training a robot to play ball

Oskar Widmark & Saam Mirghorbani

Pong in robotics

Reinforcement learning - Q-learning

- Uses Q-values to estimate the action-value for each state/action pair
- Iteratively updates values with rewards while exploring state-action space

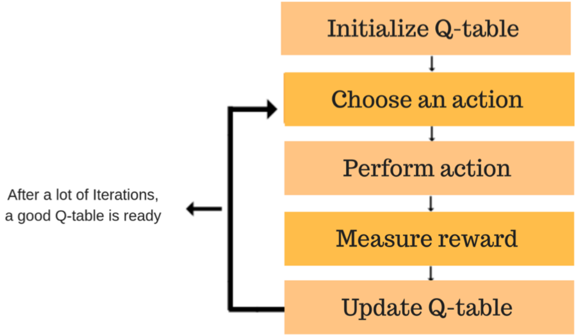


$$Q^{\pi}(s_t, a_t) = \underline{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | s_t, a_t]$$

Q-Values for the state given a particular state

Expected discounted cumulative reward

Given the state and action



Initialize Q-table

Choose an action

Perform action

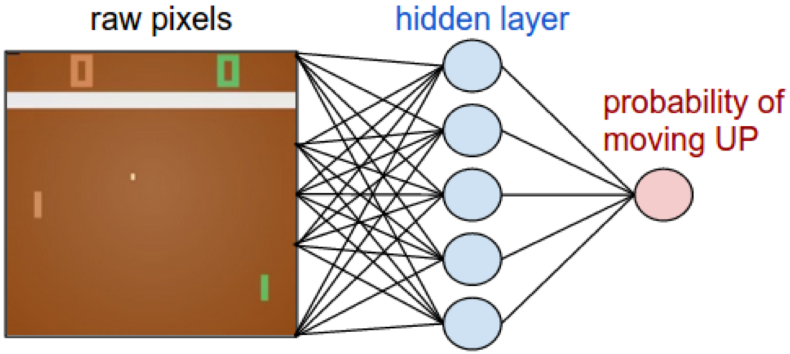
Measure reward

Update Q-table

After a lot of Iterations, a good Q-table is ready

Deep Q networks

- Combines Q-learning with deep convoultional neural networks
- Approximates a function for the optimal action that maximizes the future cumulative reward



DDPG + HER algorithm

- Deep Deterministic Policy Gradients
  - Handles continuous action spaces as opposed to DQN which can only handle discrete action spaces.
  - Suitable for robot tasks since they require continuous action spaces with multiple degrees of freedom.
- Hindsight Experience Replay
  - Lets the agent learn from mistakes.
  - Instead of receiving reward -1 for not being in target state, it will treat it as a different goal and training is simplified by letting the agent receive more rewards that differ from -1.
  - Replays each state sequence comparing different goals.
  - Combined with a neural network, the agent can learn how to achieve the original goal without even observing it during training.

HER in a single goal case

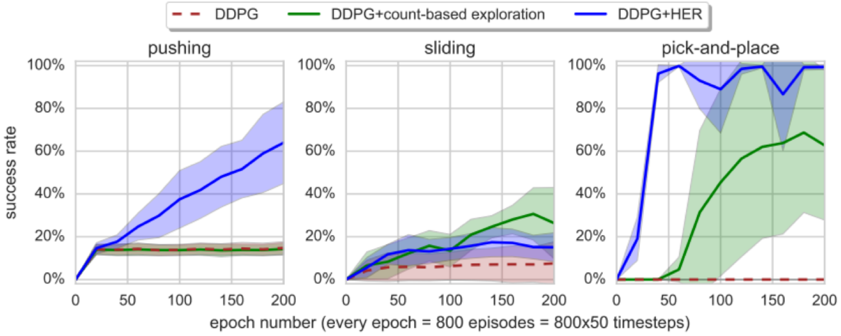
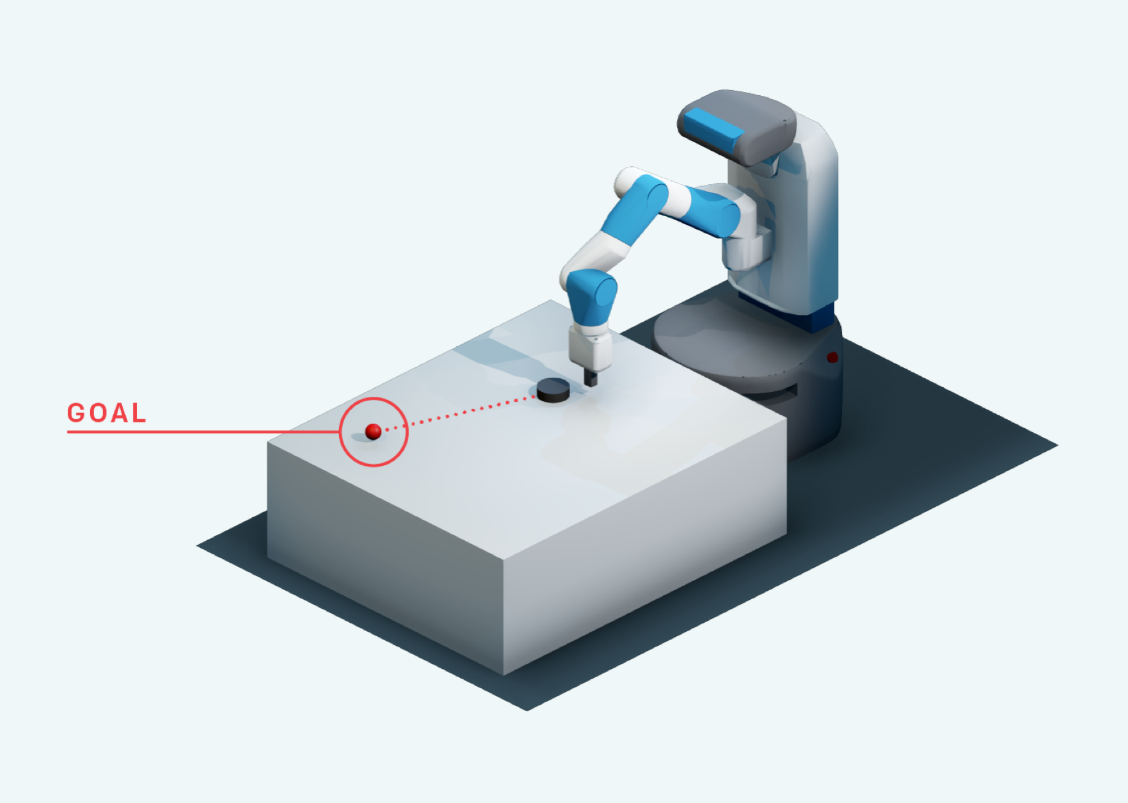


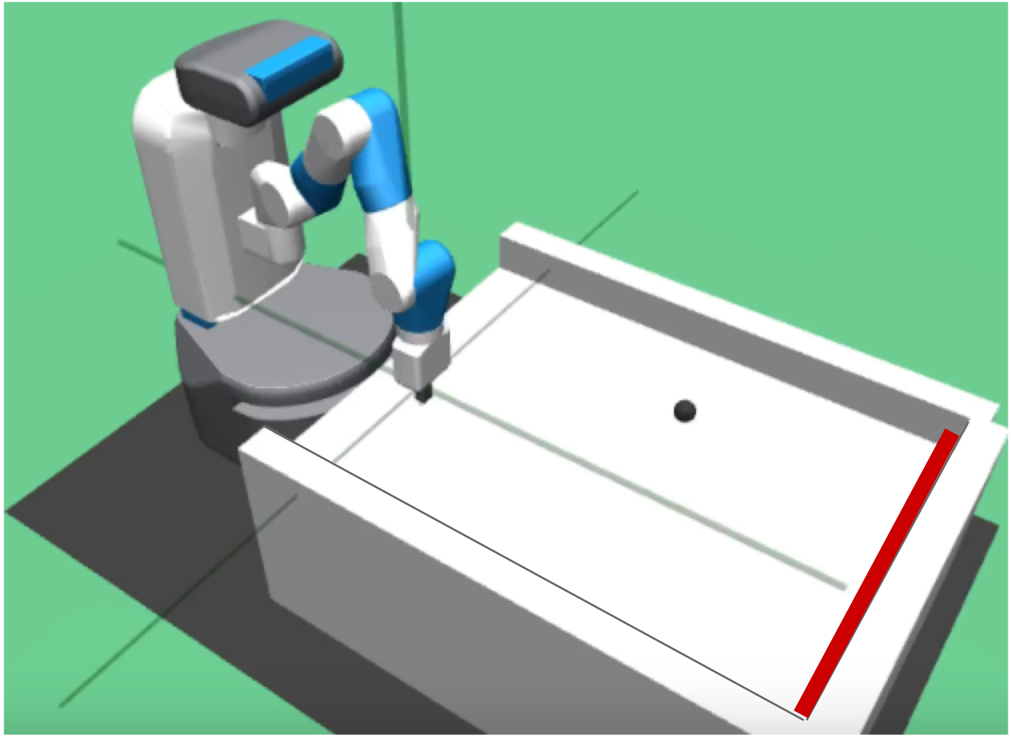
Figure 4: Learning curves for the single-goal case.

- Still better than DDPG
- Relevant for our project

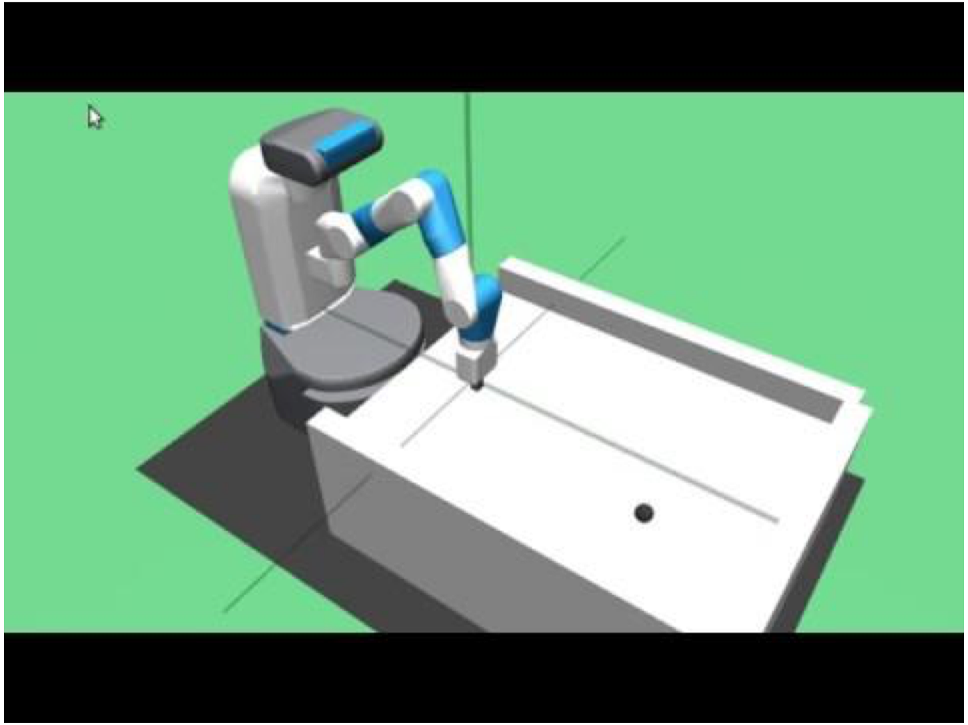
The FetchSlide environment



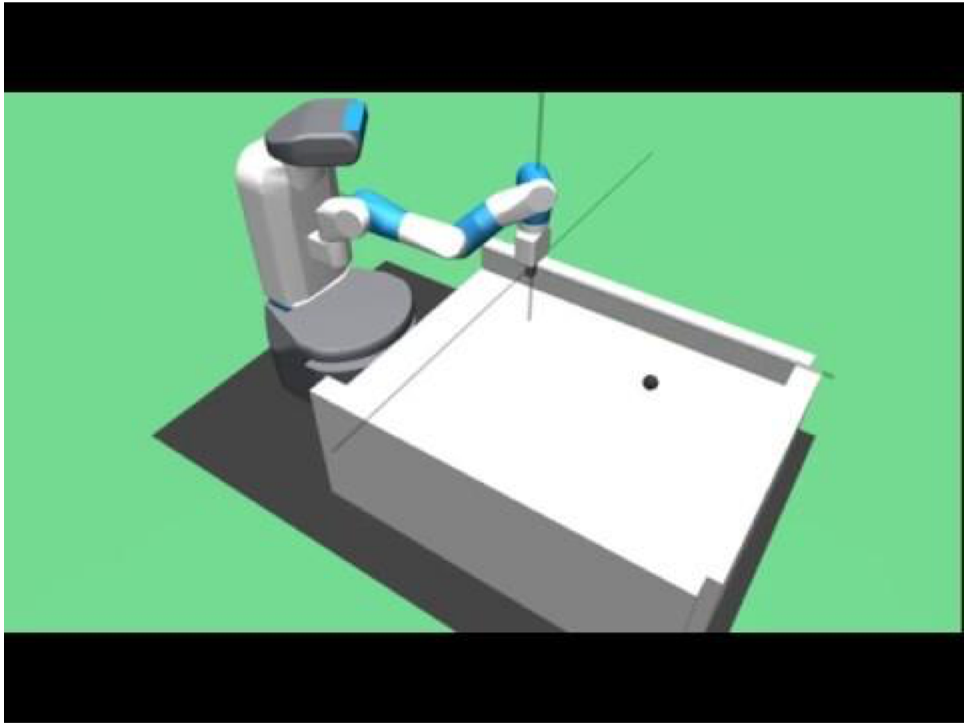
The Goalie environment



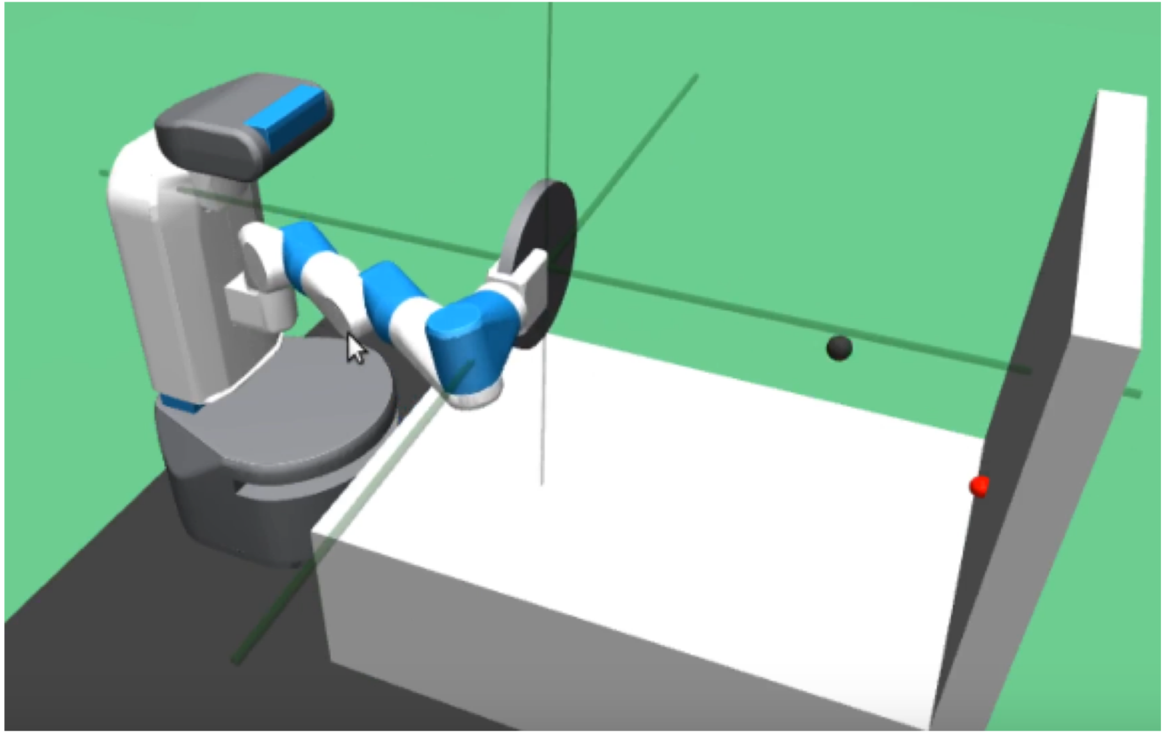
Untrained agent



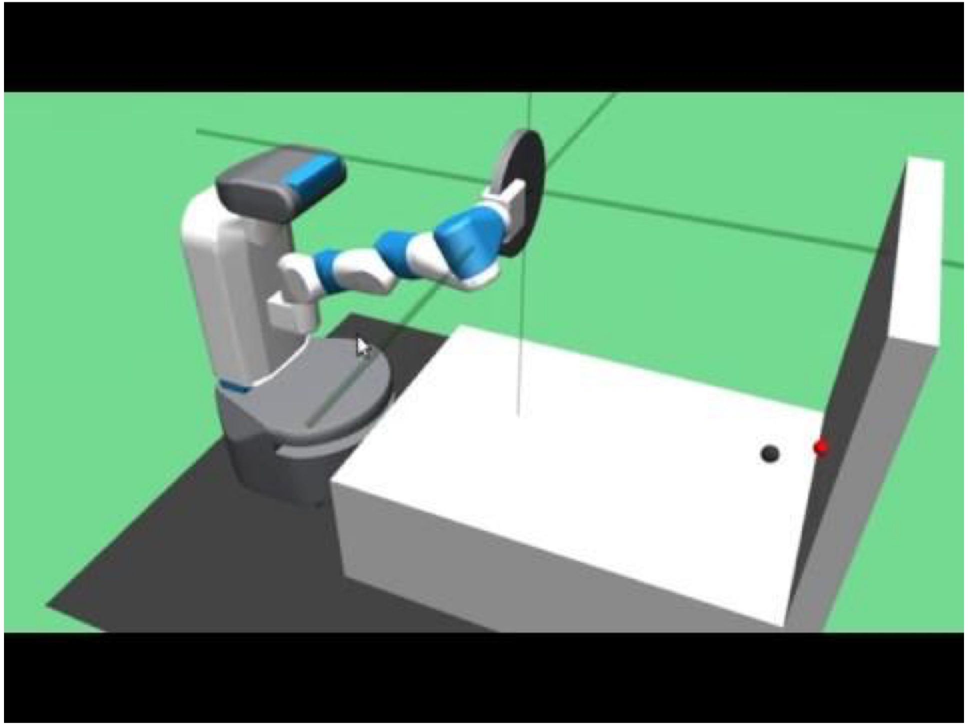
Results – Goalie



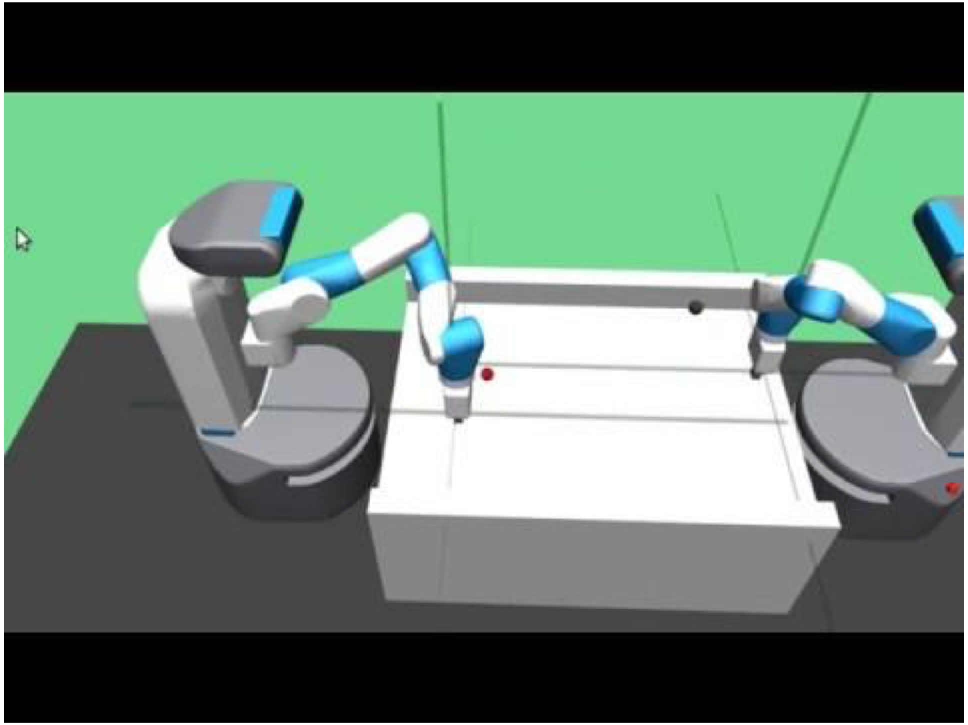
The Squash environment



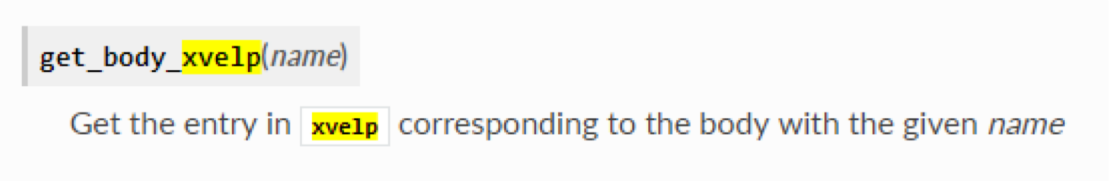
Results – Squash



Further work – Pong



???



`get_body_xvelp(name)`

Get the entry in `xvelp` corresponding to the body with the given *name*

???

Conclusions

- With carefully placed rewards, HER works quite well at learning a robot to play against itself in a Goalie environment
- Even with carefully placed rewards, HER struggles to learn a robot play in the Squash environment
- Works well in simulation environment, will probably struggle in real-life scenarios - not enough observation parameters