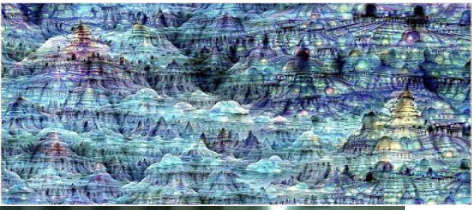
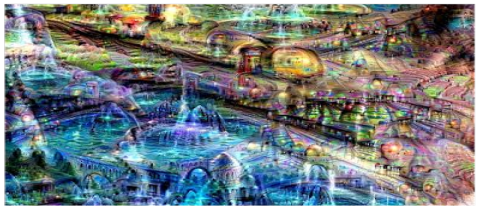
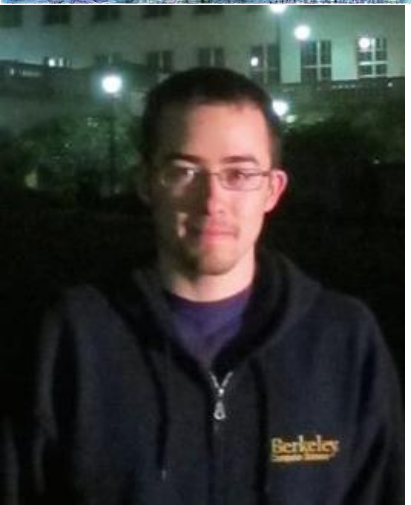


Human Learning vs. Deep Learning

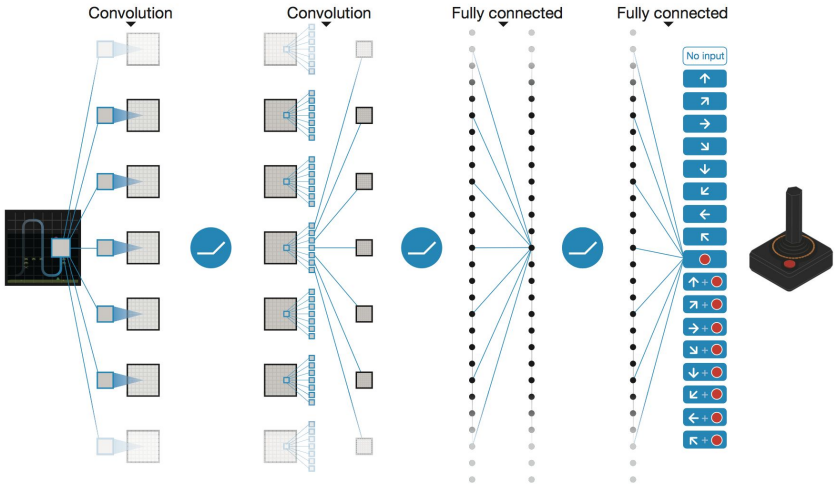
Daniel Seita

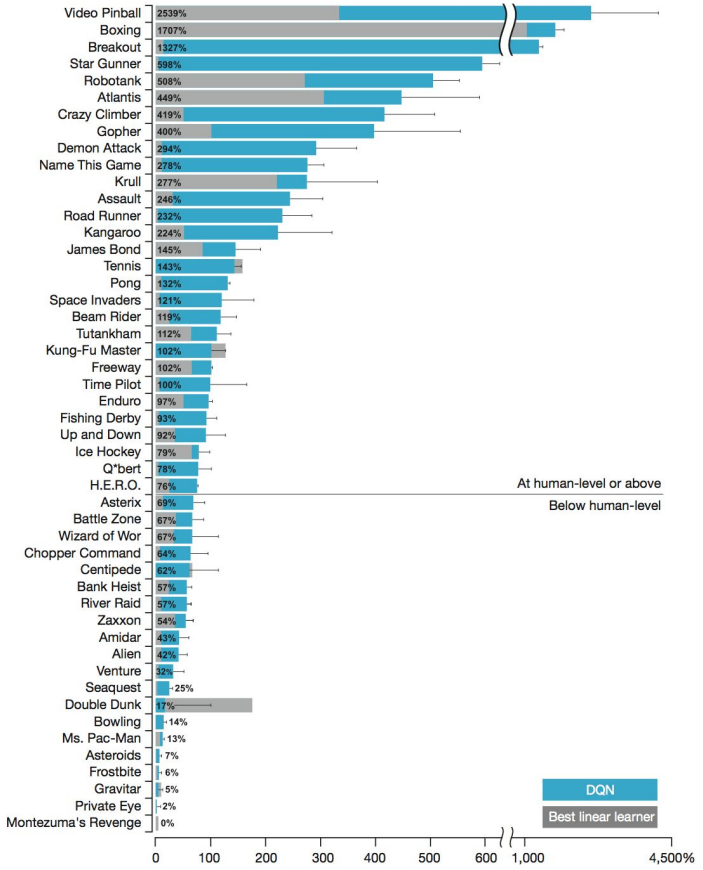



Previous Research Results

Human-Level Control Through Deep Reinforcement Learning (Nature 2015)

- Developed AI agent called “Deep Q-Network”
- Q-Learning with Q(s,a) determined by neural network
- Variants: states are (84x84x4), use experience replay, etc.
- Experiments on Atari 2600 games
  - Only pixels and game scores as input
  - Comparable to a human “professional” gamer





Experiment on Human Participants

Three Atari 2600 games:

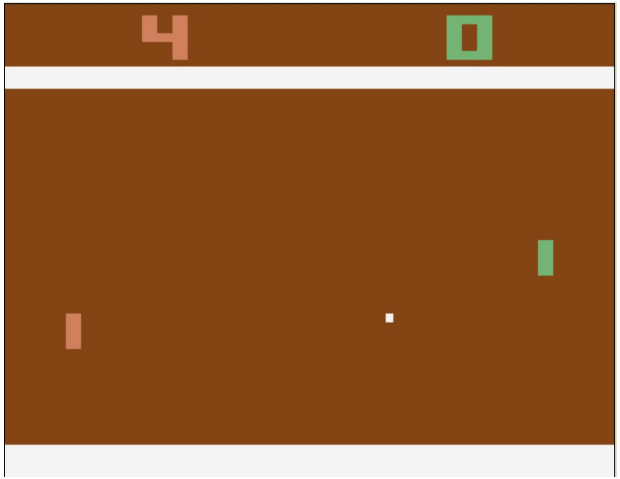
- Pong
- Breakout
- Space Invaders

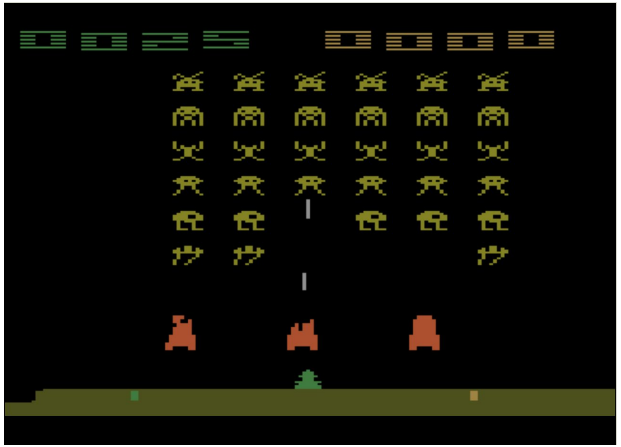
Test on Amazon Mechanical Turk:

- Tiers ($1.00, $1.50, $2.00)
- Exactly 10 minutes per game
- Record their game logs
- Different game ordering

Example questions:

- State prior gaming experience
- Did you play well? (Likert)
- What was your final strategy?





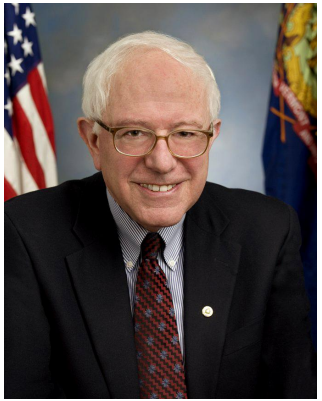
Some Current Results

Observations after 28 players (out of 120 planned)

- Some skilled players with Space Invaders
- Pong and Breakout were challenging
- Payment thresholds probably too low

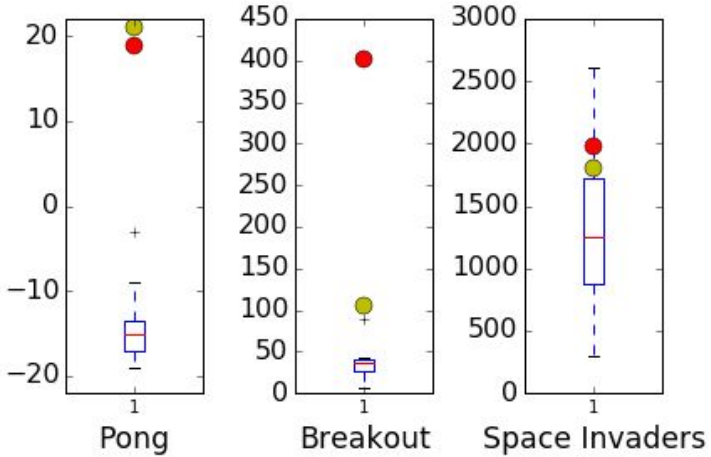
Player feedback

- Many didn't like game controls (arrow keys)
- Some didn't know to use certain controls
- Some said experience in other games helped



Quantitative Results

- Box-plots of human **high** scores
- Red = DQN **average** score
- Yellow = my **average** score



Some Player Comments (Not Edited) and Future Work

After Pong: *My strategy during this game was an utter disaster throughout the entire phase of play.*

Another Pong comment: *Breakout helped learn how to hit the ball at an angle*

After Space Invaders: *Eliminate most off the lowest aliens first, then concentrate on an outer column*

After all games: *If there were a way to have better control of the paddle, I would try to hit more angular shots in the PONG game. The Atari paddle controller on the old 2600 system made this easily possible. The keyboard controls show weakness [...]*

Future work for this week:

- Run DQN algorithm, except use human data for initial exploration policy
- Run Q-Learning on this data to see how much worse it is than results from Nature letter