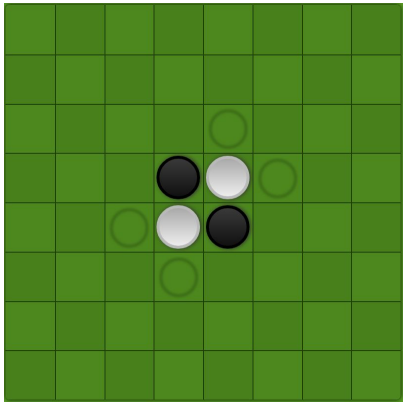
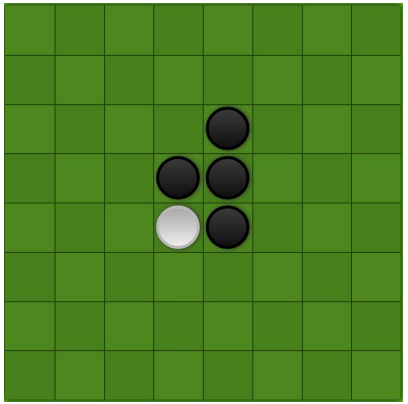
Curriculum Learning for Reversi

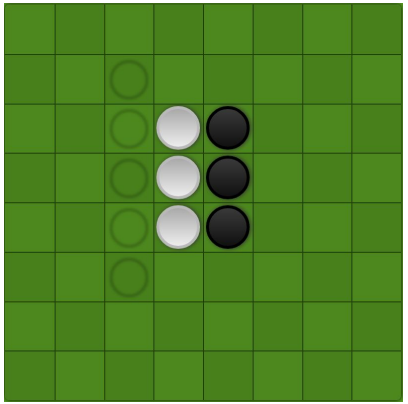
Sasha Fedchin
Erli Cai

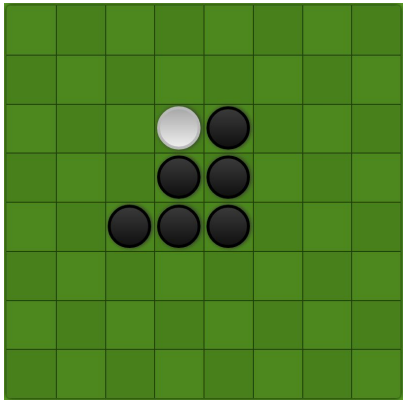
Reversi



- Simple rules
- Used in research
- Large state complexity
- Reward at the end
- Afterstates







N-step SARSA with afterstates

$$G_{t:t+n} \doteq R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{n-1} R_{t+n} + \gamma^n Q_{t+n-1}(S_{t+n}, A_{t+n}), \quad n \geq 1, 0 \leq t < T - n,$$

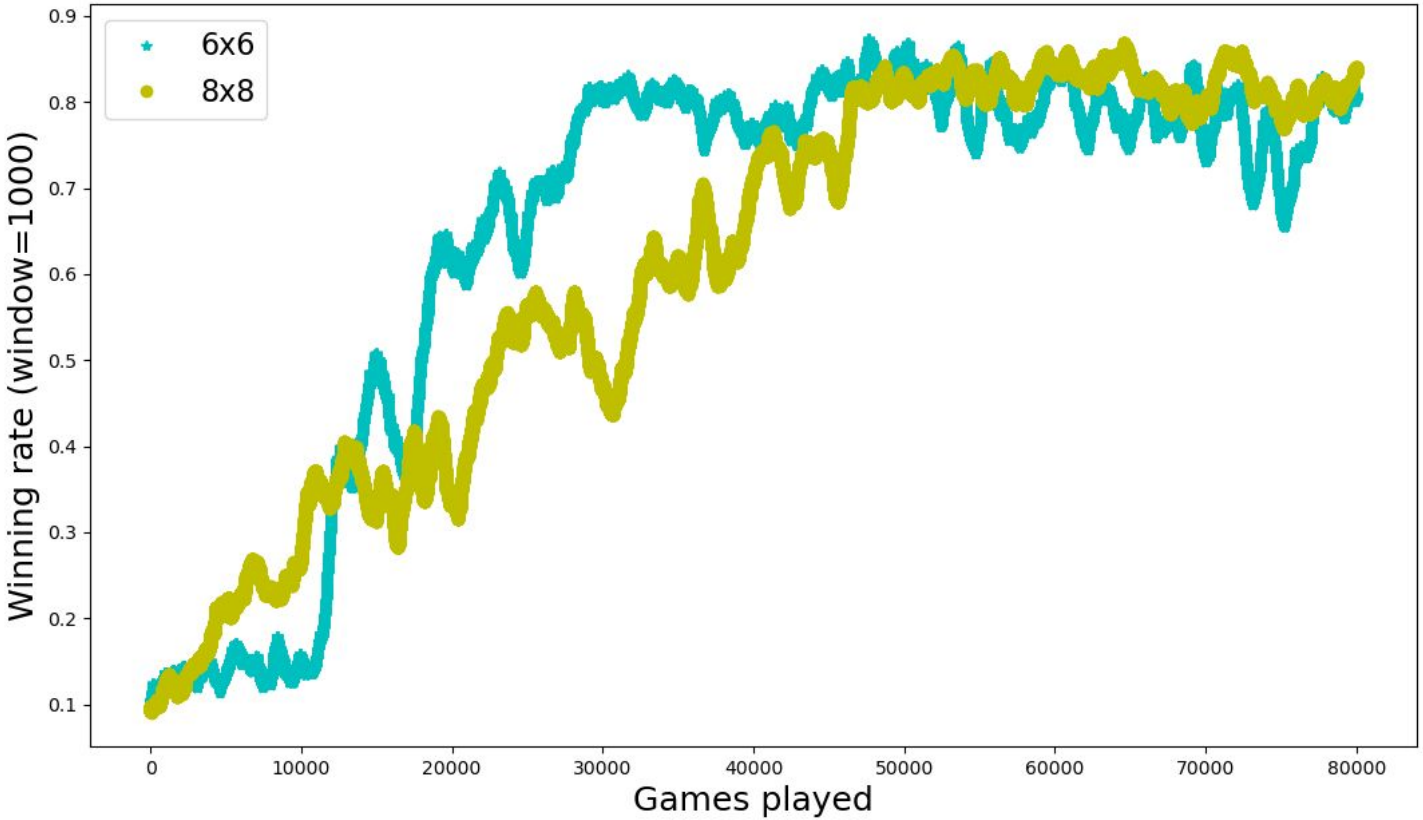
$$G_{t:t+n} = R_{t+1} + \gamma R_{t+2} + .. + \gamma^{n-1} R_{t+n} + \gamma^n Q_{t+n-1}(S_{t+n})$$

$$R_T = 0 \text{ for all T except terminal state}$$

$$Q_{t+n}(S_t, A_t) \doteq Q_{t+n-1}(S_t, A_t) + \alpha \left[G_{t:t+n} - Q_{t+n-1}(S_t, A_t) \right]$$

$$Q_{t+n}(S_t) = Q_{t+n-1}(S_t) + \alpha [G_{t:t+n} - Q_{t+n-1}(S_t)]$$

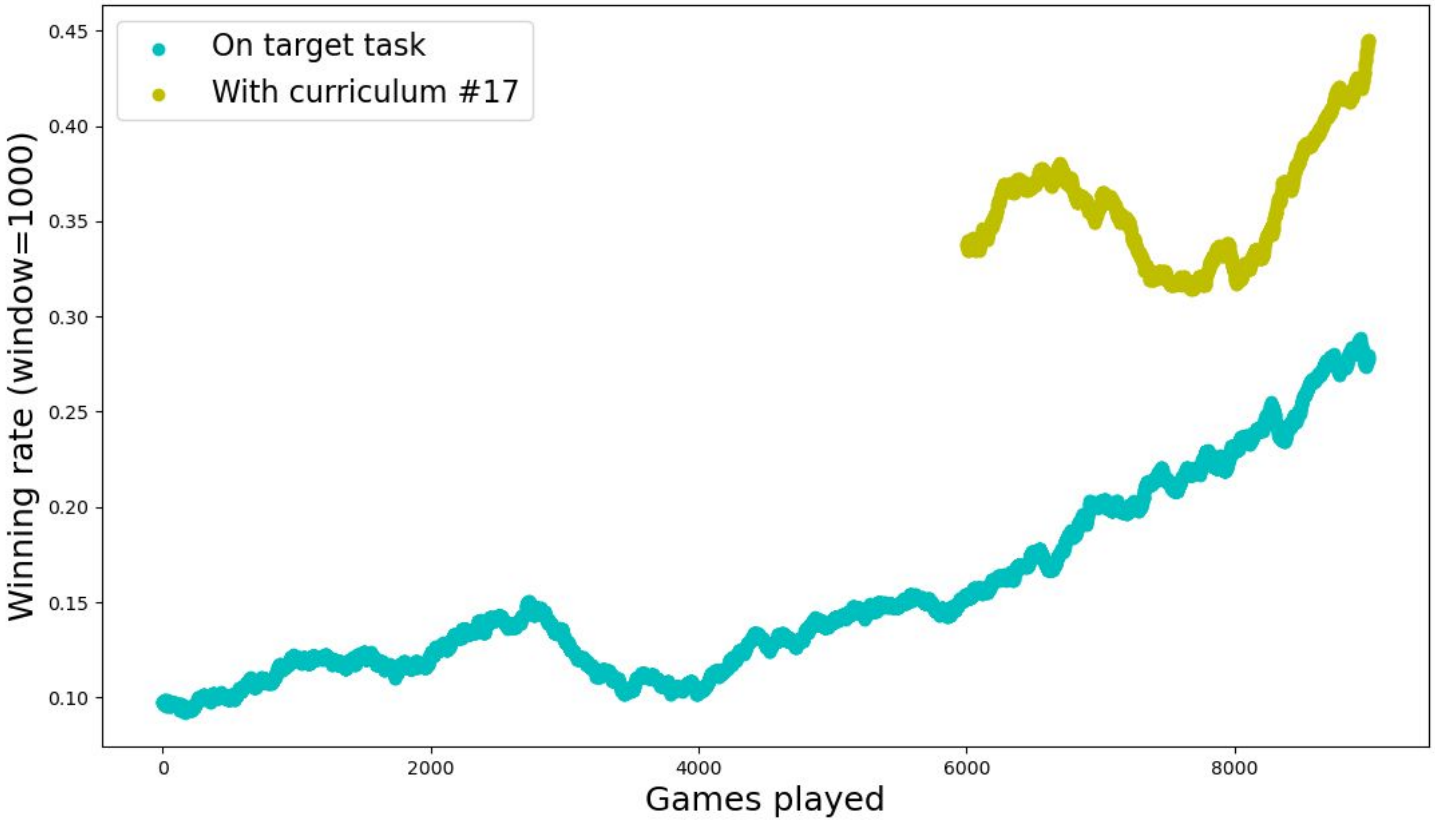
Training on the target task



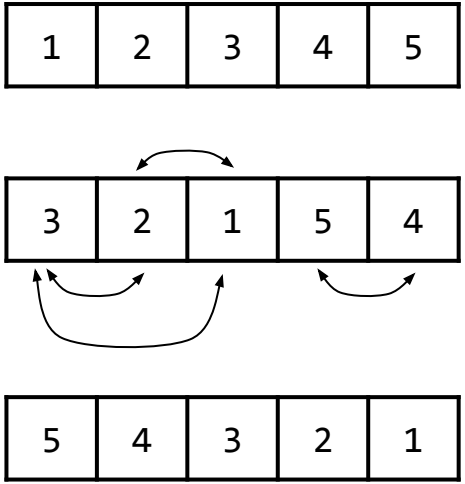
Curriculum learning in Reversi

- Tasks (opponent, depth, epsilon)
- 20 random curricula 2 tasks total, 3rd task is target
- Rank curricula for 6x6 and 8x8 boards
- Hypothesis: rankings are similar

Curriculum vs No Curriculum



Kendall tau coefficient



Example ranking, 10 pairs total

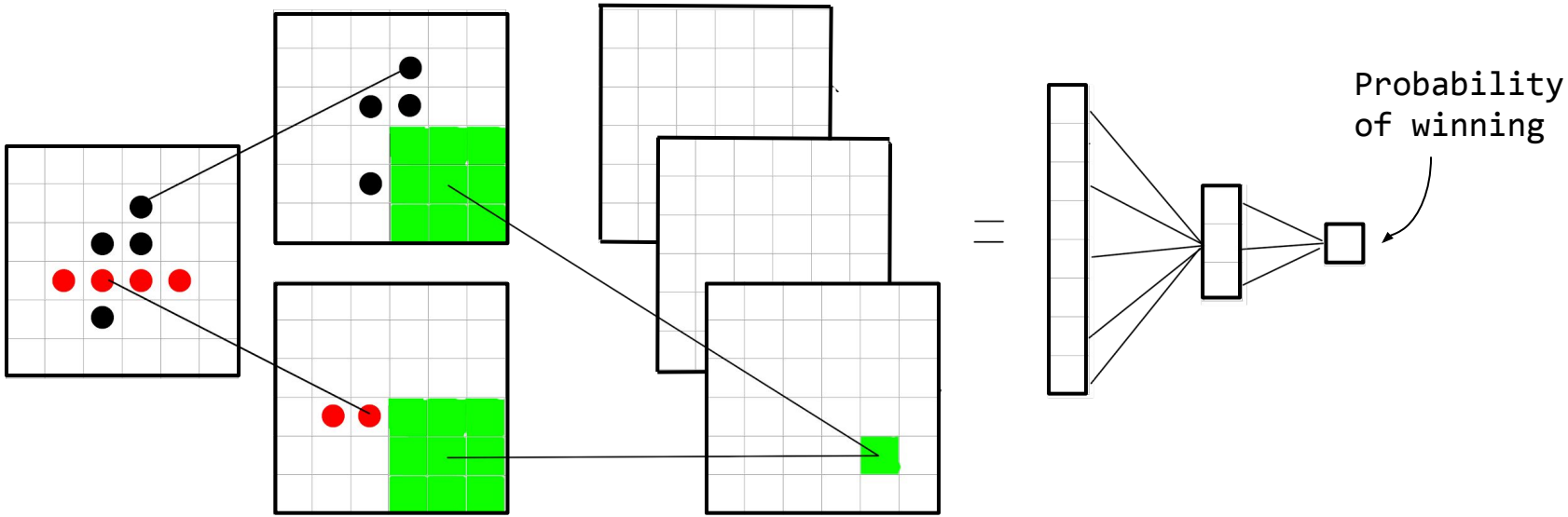
4 discordant, 6 concordant, $\tau = (6-4)/10 = 0.2$

10 discordant, 0 concordant, $\tau = (0-10)/10 = -1$

Our case: $\tau = -0.17$, $p = 0.36$

Convolutional Neural Networks

- Kernel of 3 is best
- Larger board -> more channels and layers
- One CNN to approximate values of all states



Thank you!