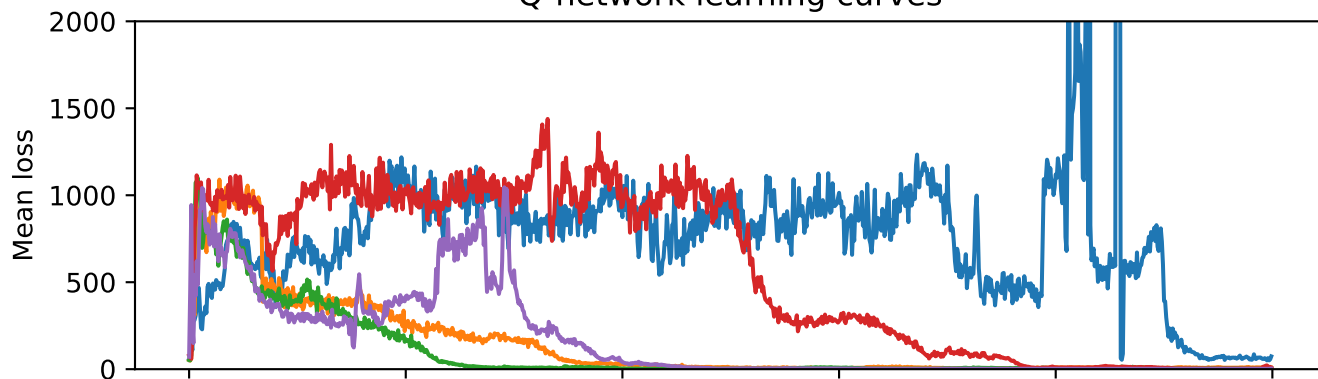
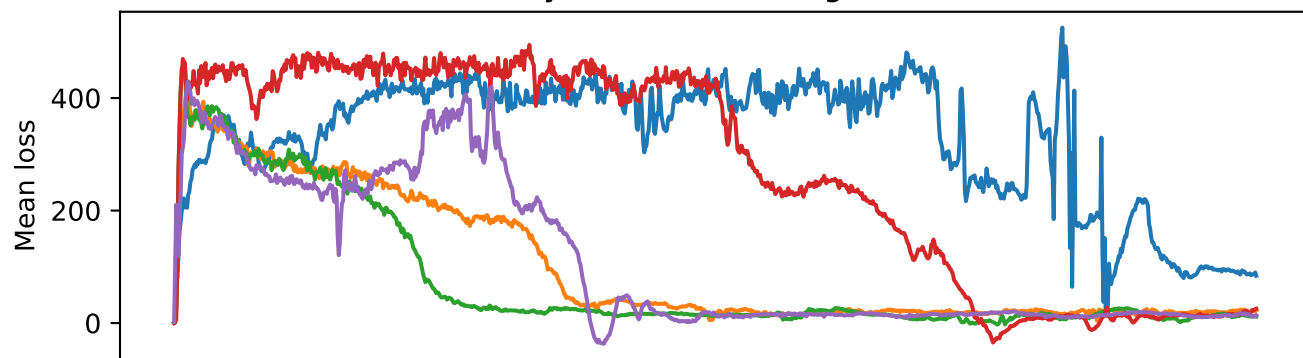


Q-network learning curves



Policy network learning curves



Smoothed returns (moving average window = 50)

