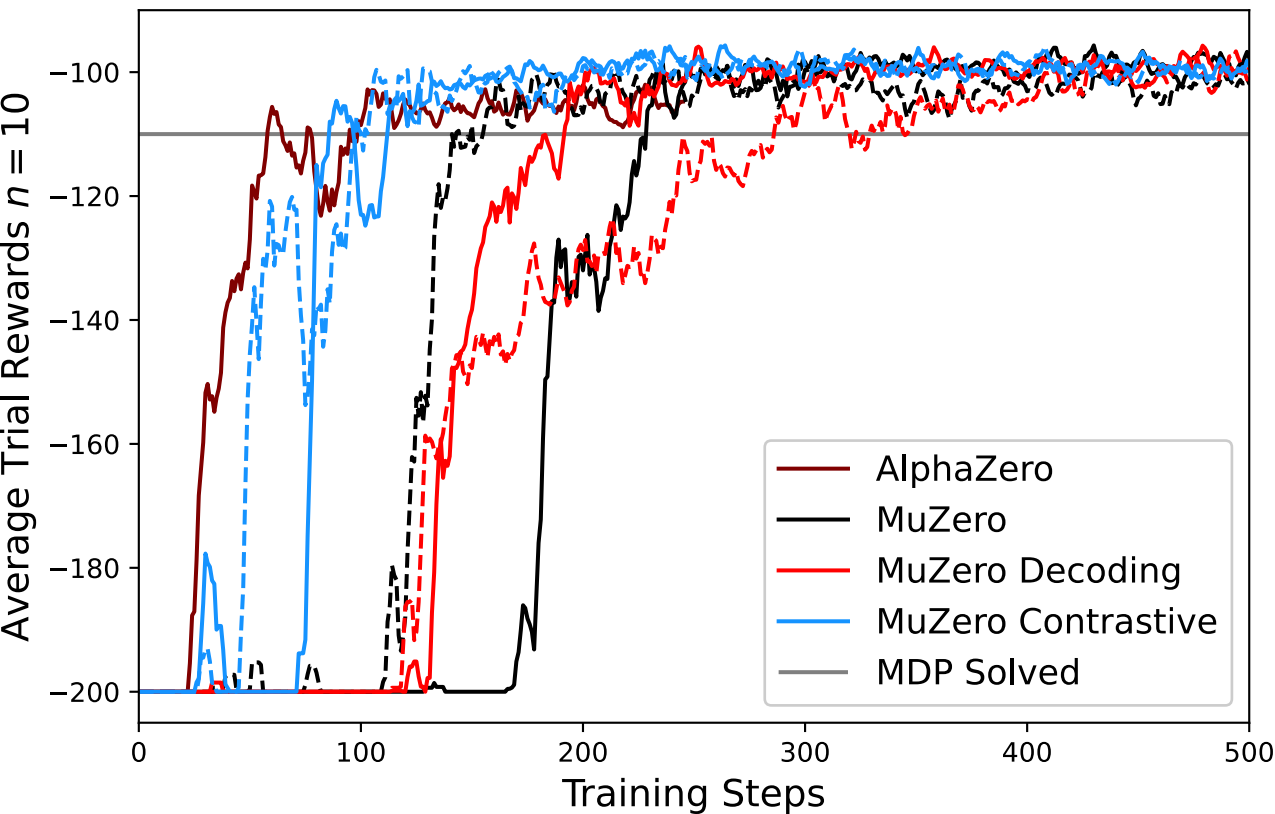MountainCar Mean Trial Rewards (solid $L = 8$; dashed $L = 4$)

During Training

"Blindfolded" Control of Trained Agents

Average Trial Rewards $n = 10$

Training Steps

Observation Sparsity (Refresh Frequency)

AlphaZero
MuZero
MuZero Decoding
MuZero Contrastive
MDP Solved