Model Average Reward over Training vs Baselines SAC 2000 PPO (continuous) DQN 1750 Avg. Max. Available 1500 Episode Average Reward 1250 1000 750 500 250 0 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 1e6 Training Iterations (Millions)