Bayesian Q-learning **Greedy selection** Q-value sampling Myopic VPI selection Reward Number of episodes