

# 1 Deterministic Environment: Exploration Rate (Q-learning)

## 1.1 Epsilon: 0.1

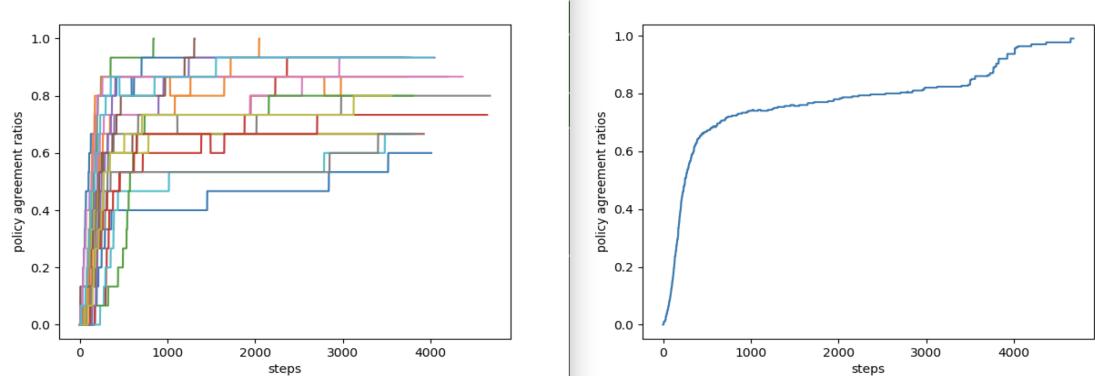


Figure 1: Deterministic Environment - Epsilon: 0.1 (Q-learning)

## 1.2 Epsilon: 0.05

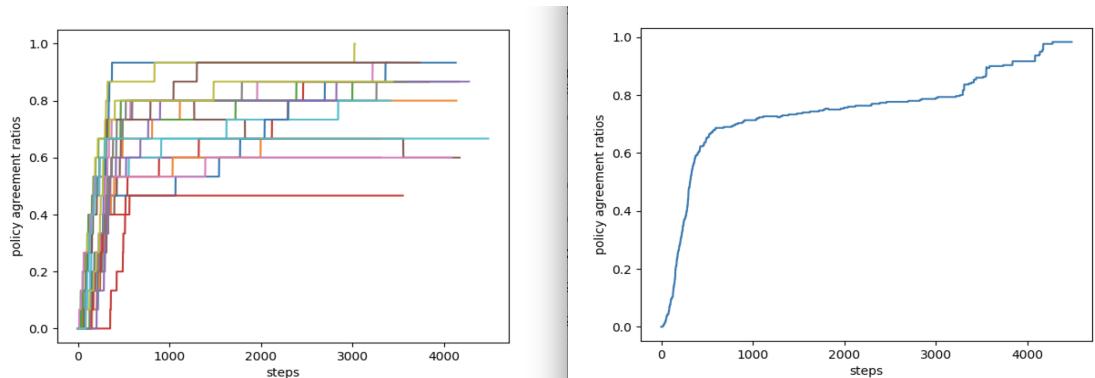


Figure 2: Deterministic Environment - Epsilon: 0.05 (Q-learning)

## 1.3 Comparison

### 2 TAMER using epsilon-greedy

### 3 Noise: 0.1

### 4 Noise: 0.5

### 5 Noise: 0.9

### 6 Different Learning Rate with Noise 0.2

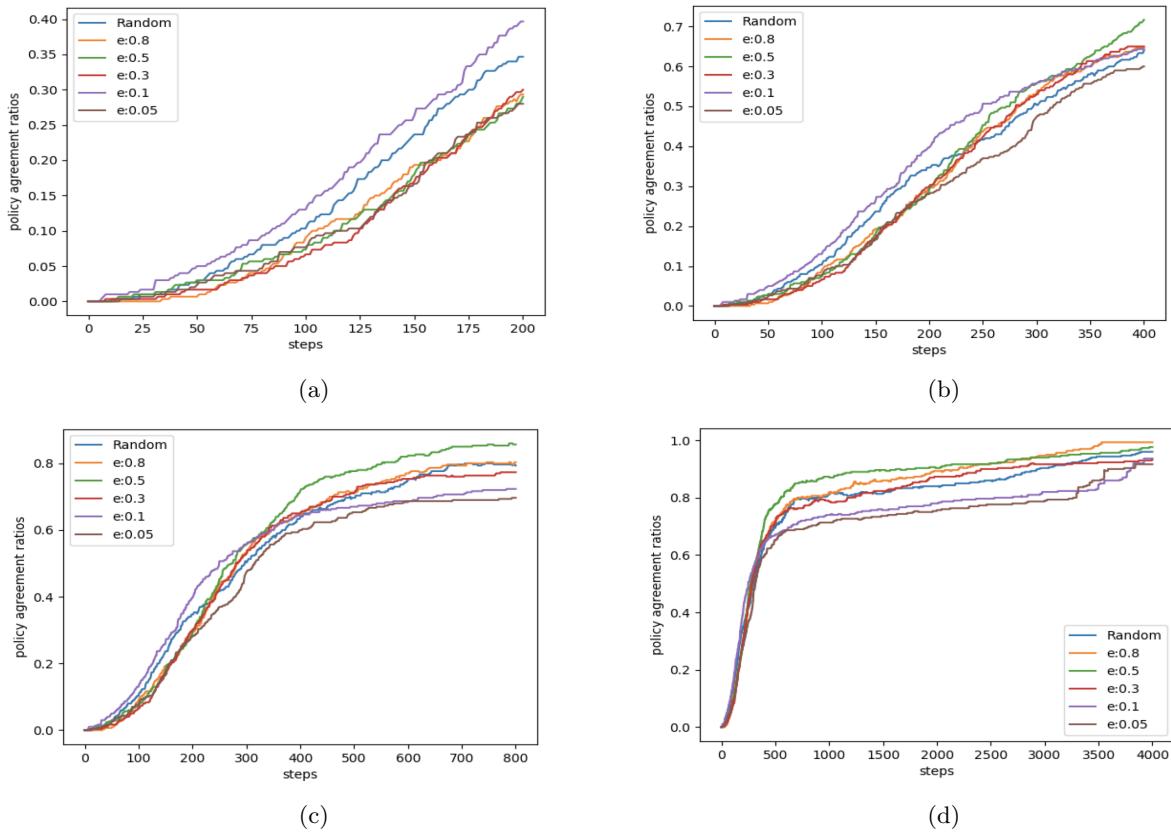


Figure 3: Average policy agreement ratios at each time step of experiments with different epsilon values. (Q-learning)

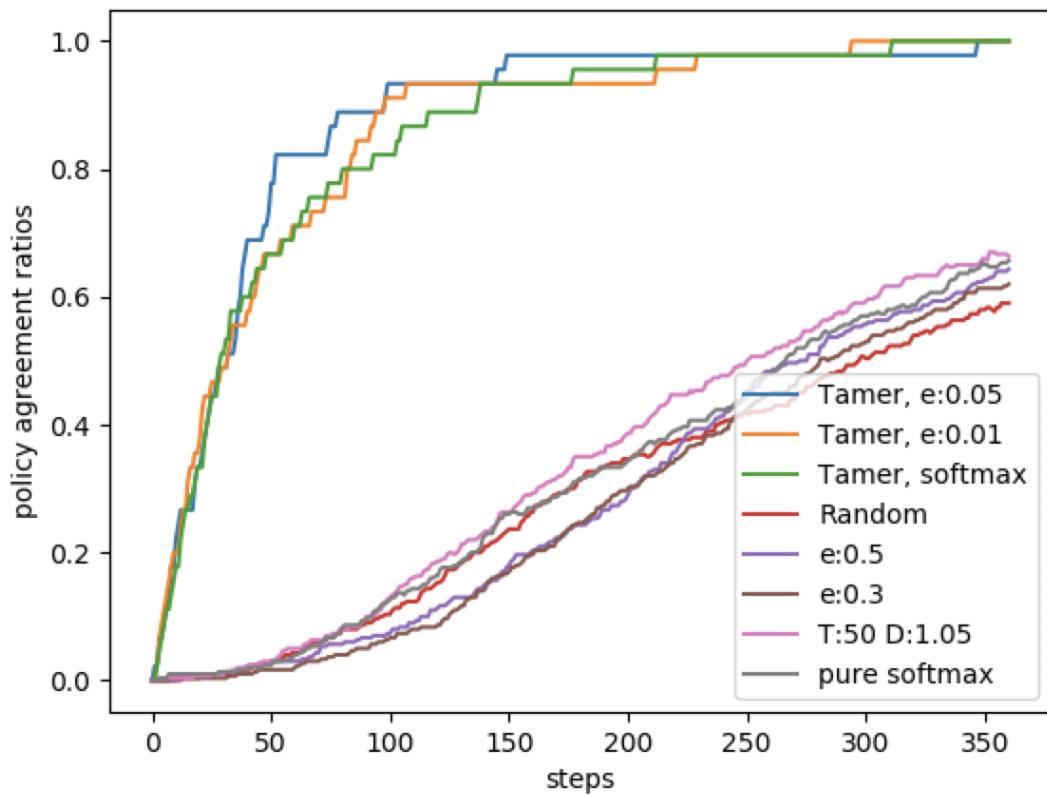


Figure 4: TAMER using epsilon-greedy. (TAMER)

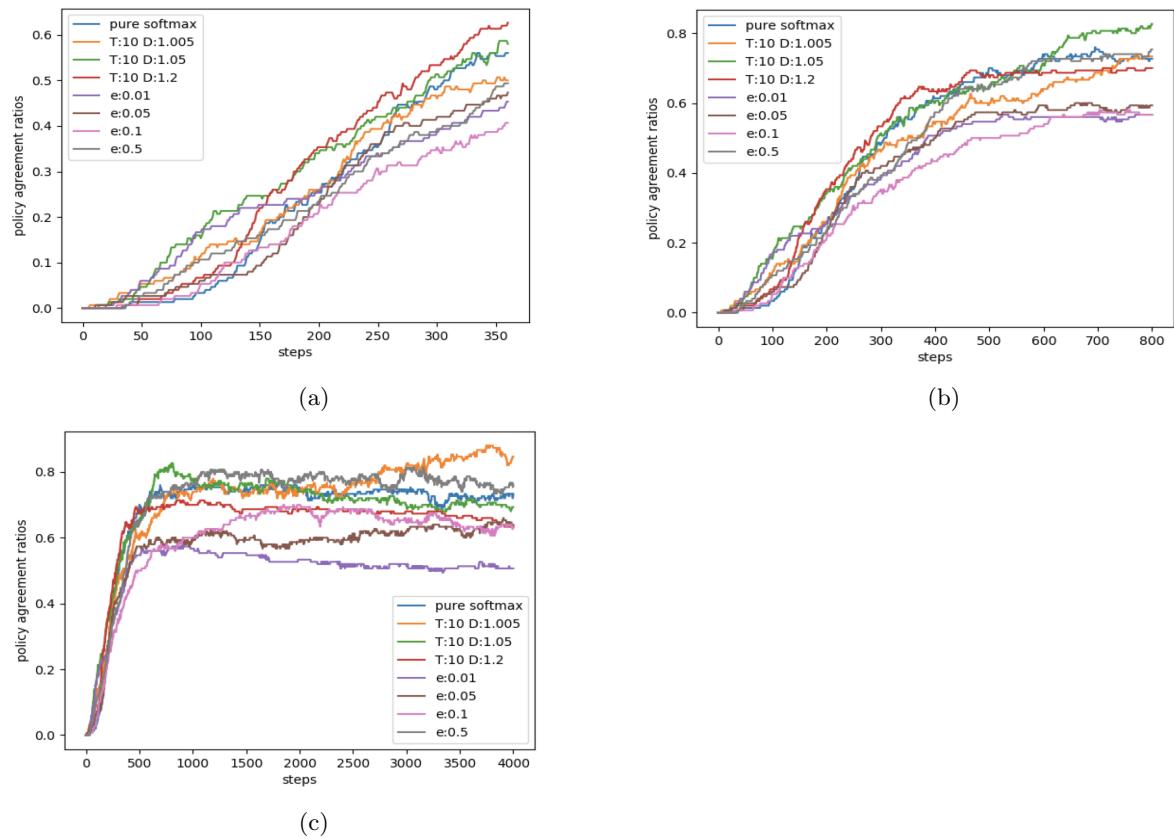
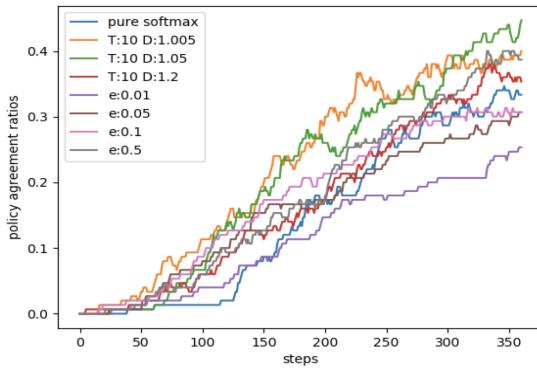
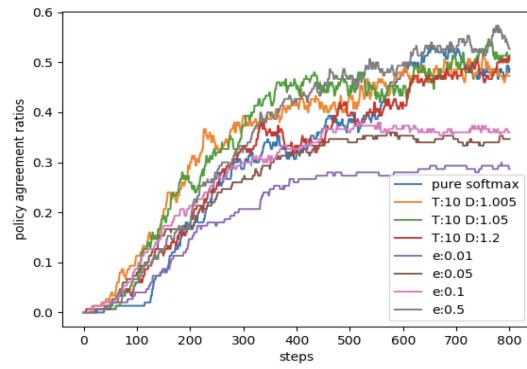


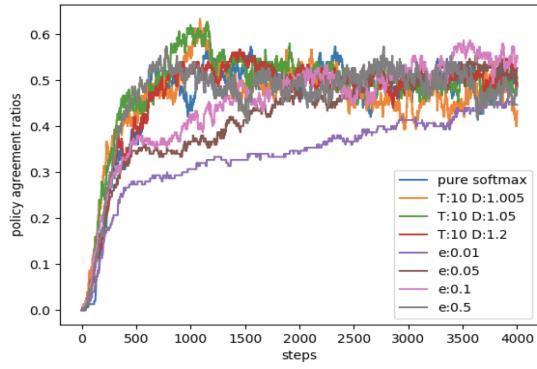
Figure 5: Noise: 0.1



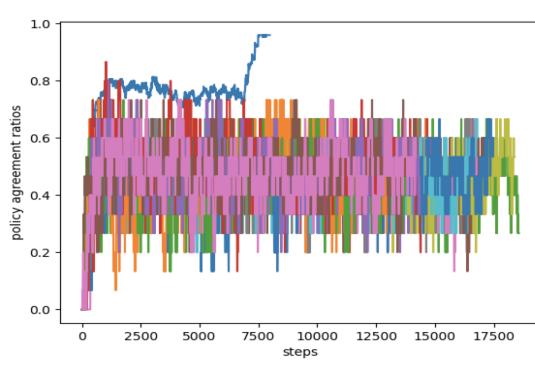
(a)



(b)



(c)



(d)

Figure 6: Noise: 0.5

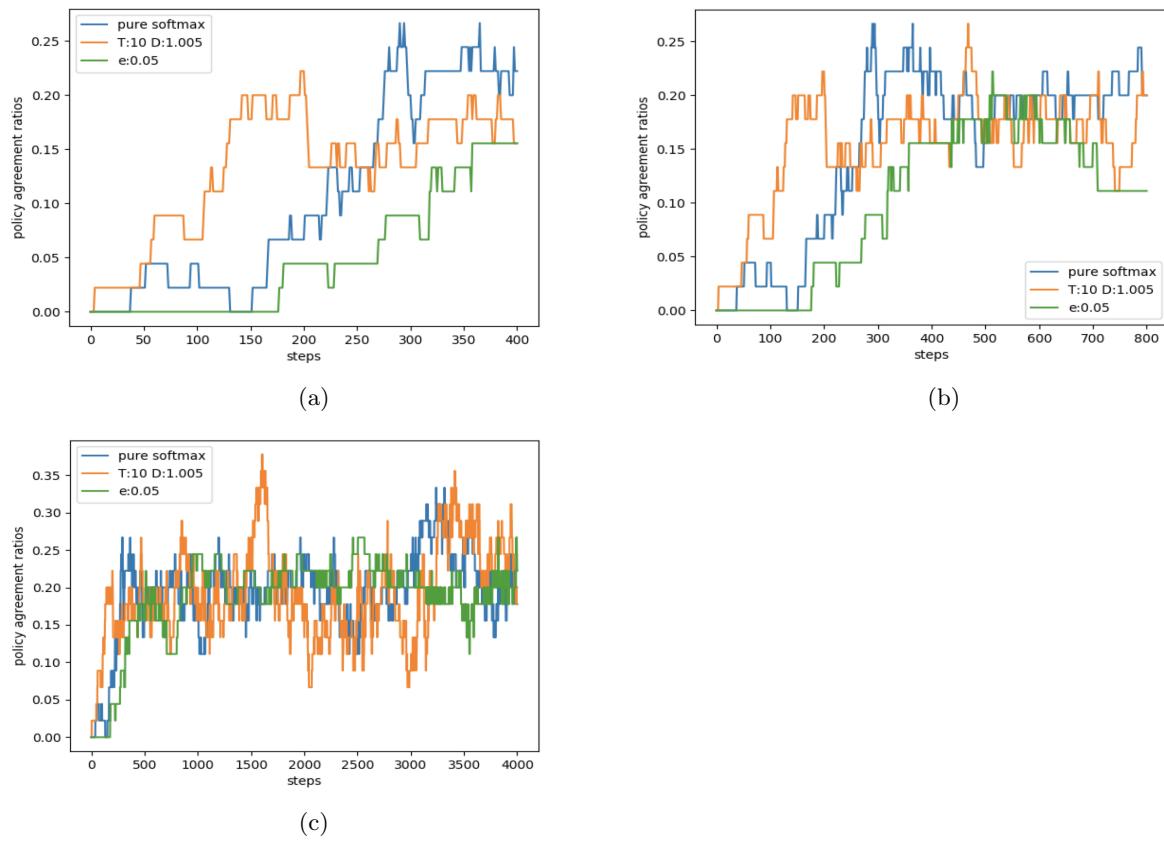


Figure 7: Noise: 0.9

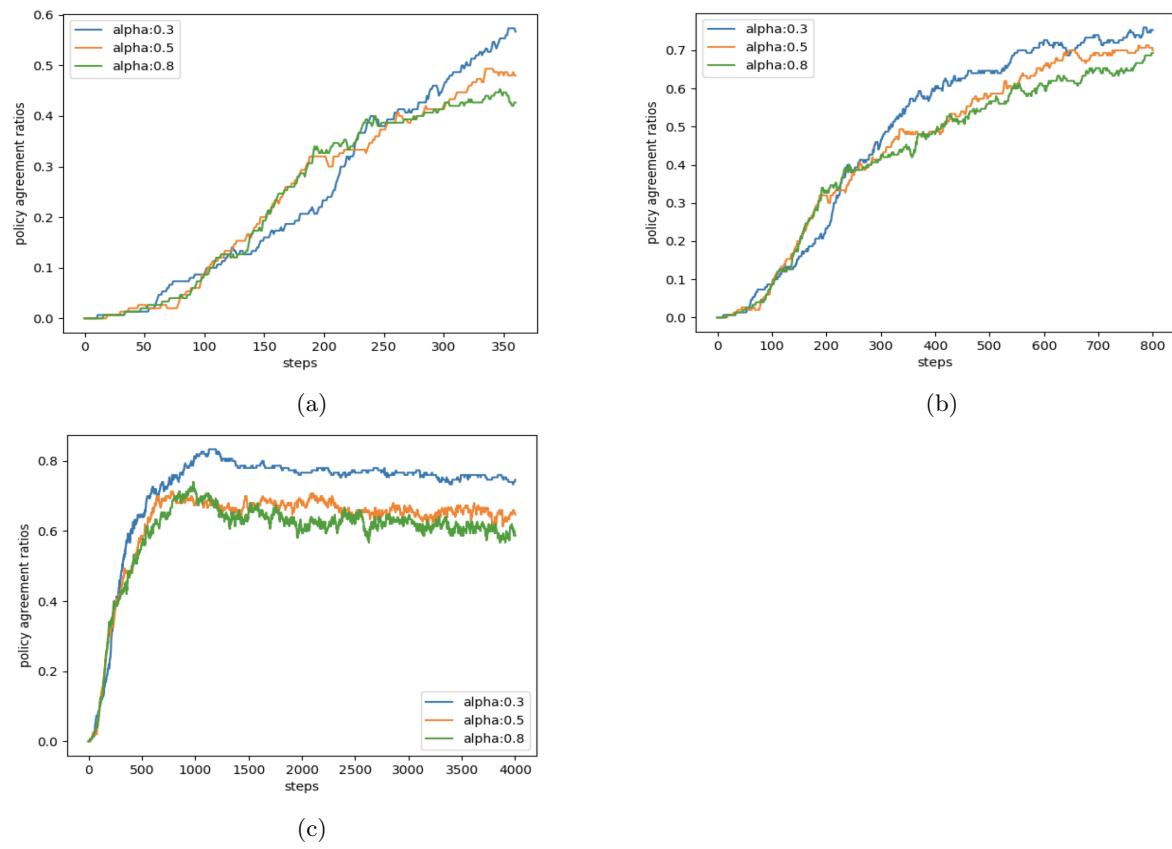


Figure 8: Different Learning Rate with Noise 0.2 (Q-Learning)