

- (1) Init $Q_{s,a}$ tables
- (2) Get (s, a, r, s') from the environment
- (3) $Q_{s,a} = (1 - \alpha)Q_{s,a} + \alpha(r + \gamma \max_{a' \in A} Q_{s',a'})$
- (4) Check convergence