$Q(s, a)$ after each episode; Q_* down: 0.81, 0.90, 1.00, up: 0.00, 0.73, 0.81

