Blackjack - results

All policies were tested on 10000 random episodes.

1 Deterministic optimal policy

```
####################
Deterministic policy: GameStats(wins=0.4298, draws=0.0962, losses=0.474)
```

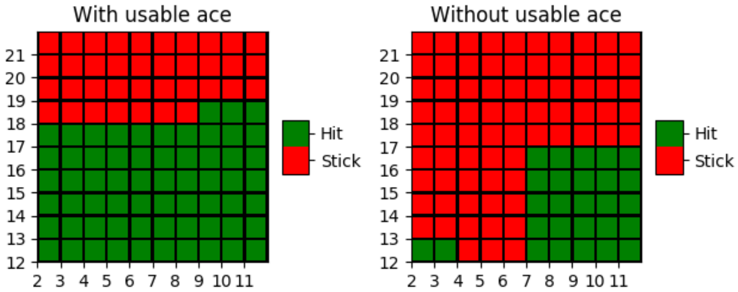


Figure 1: Deterministic policy

2 Monte Carlo Exploring States

```
####################
MonteCarloExploringStates (50000 epochs, 4.62s):
    GameStats(wins=0.4291, draws=0.096, losses=0.4749)
MonteCarloExploringStates (500000 epochs, 47.48s):
    GameStats(wins=0.4327, draws=0.0889, losses=0.4784)
```

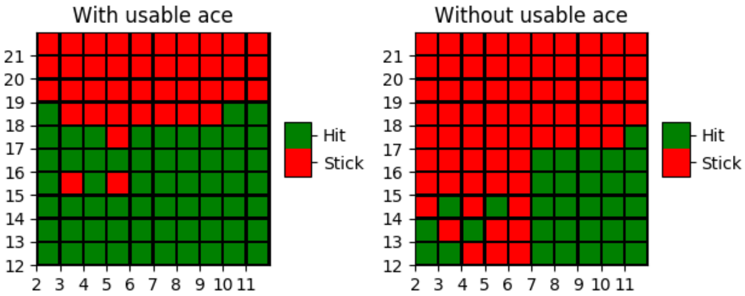


Figure 2: Policy after 50k episodes.

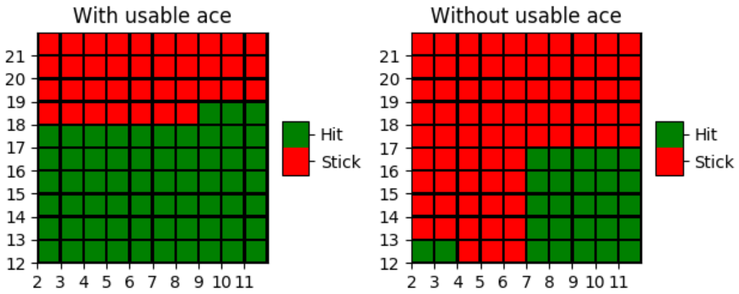


Figure 3: Policy after 500k episodes.

3 On-policy first visit Monte Carl control

```
####################
MonteCarloOnPolicyFirstVisit (50000 epochs, 5.48s):
    GameStats(wins=0.3879, draws=0.078, losses=0.5341) eps: 0.2
MonteCarloOnPolicyFirstVisit (50000 epochs, 5.78s):
    GameStats(wins=0.402, draws=0.0868, losses=0.5112) eps: 0.1
MonteCarloOnPolicyFirstVisit (50000 epochs, 5.55s):
    GameStats(wins=0.4109, draws=0.0784, losses=0.5107) eps: 0.05
MonteCarloOnPolicyFirstVisit (500000 epochs, 57.42s):
    GameStats(wins=0.3995, draws=0.0752, losses=0.5253) eps: 0.2
MonteCarloOnPolicyFirstVisit (500000 epochs, 58.44s):
    GameStats(wins=0.4169, draws=0.0862, losses=0.4969) eps: 0.1
MonteCarloOnPolicyFirstVisit (500000 epochs, 57.97s):
    GameStats(wins=0.4161, draws=0.0848, losses=0.4991) eps: 0.05
```

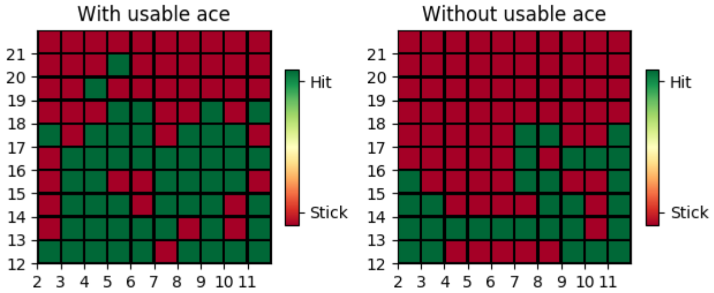


Figure 4: Policy after 50k episodes for $\epsilon = 0.05$ (5% chance of taking action with worse value)

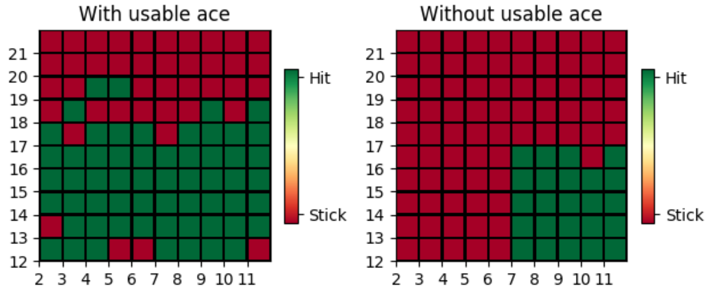


Figure 5: Policy after 500k episodes for $\epsilon = 0.1$ (10% chance of taking action with worse value)

4 Sarsa

To obtain exploring policy while keeping it updated to the computed values, the worse action was taken with probability $\frac{1}{round}$, where *round* is the number of episodes taken already.

```
####################
Sarsa (50000 epochs, 5.40s):
    GameStats(wins=0.4321, draws=0.0806, losses=0.4873) alpha: 0.01 gamma: 0.5
Sarsa (50000 epochs, 5.55s):
    GameStats(wins=0.4223, draws=0.0835, losses=0.4942) alpha: 0.01 gamma: 0.9
Sarsa (50000 epochs, 5.21s):
    GameStats(wins=0.4234, draws=0.0911, losses=0.4855) alpha: 0.01 gamma: 0.99
Sarsa (50000 epochs, 5.43s):
    GameStats(wins=0.4154, draws=0.0917, losses=0.4929) alpha: 0.1 gamma: 0.5
Sarsa (50000 epochs, 5.33s):
    GameStats(wins=0.4121, draws=0.0984, losses=0.4895) alpha: 0.1 gamma: 0.9
Sarsa (50000 epochs, 5.31s):
    GameStats(wins=0.4182, draws=0.0977, losses=0.4841) alpha: 0.1 gamma: 0.99
Sarsa (50000 epochs, 5.72s):
    GameStats(wins=0.4074, draws=0.0958, losses=0.4968) alpha: 0.5 gamma: 0.5
Sarsa (50000 epochs, 5.51s):
    GameStats(wins=0.4122, draws=0.099, losses=0.4888) alpha: 0.5 gamma: 0.9
Sarsa (50000 epochs, 5.52s):
    GameStats(wins=0.4091, draws=0.0949, losses=0.496) alpha: 0.5 gamma: 0.99
Sarsa (500000 epochs, 53.39s):
    GameStats(wins=0.4176, draws=0.0971, losses=0.4853) alpha: 0.01 gamma: 0.5
Sarsa (500000 epochs, 54.85s):
    GameStats(wins=0.4235, draws=0.0992, losses=0.4773) alpha: 0.01 gamma: 0.9
Sarsa (500000 epochs, 55.27s):
    GameStats(wins=0.4294, draws=0.0921, losses=0.4785) alpha: 0.01 gamma: 0.99
Sarsa (500000 epochs, 56.85s):
    GameStats(wins=0.4259, draws=0.0933, losses=0.4808) alpha: 0.1 gamma: 0.5
Sarsa (500000 epochs, 58.65s):
    GameStats(wins=0.4192, draws=0.1013, losses=0.4795) alpha: 0.1 gamma: 0.9
Sarsa (500000 epochs, 61.13s):
    GameStats(wins=0.4186, draws=0.0988, losses=0.4826) alpha: 0.1 gamma: 0.99
Sarsa (500000 epochs, 65.47s):
    GameStats(wins=0.4159, draws=0.0915, losses=0.4926) alpha: 0.5 gamma: 0.5
Sarsa (500000 epochs, 63.24s):
    GameStats(wins=0.4129, draws=0.0988, losses=0.4883) alpha: 0.5 gamma: 0.9
Sarsa (500000 epochs, 60.61s):
    GameStats(wins=0.4069, draws=0.0965, losses=0.4966) alpha: 0.5 gamma: 0.99
```

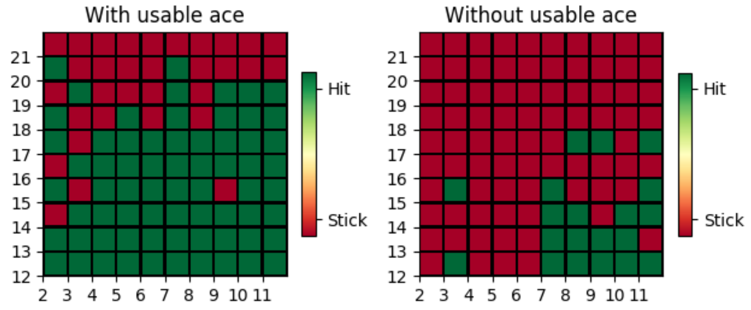


Figure 6: Policy after 50k episodes for $\alpha = 0.01$, $\gamma = 0.5$ ($\frac{1}{50000}$ chance of taking action with worse value)

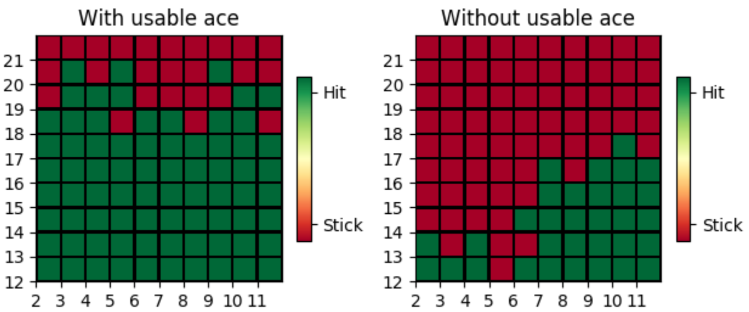
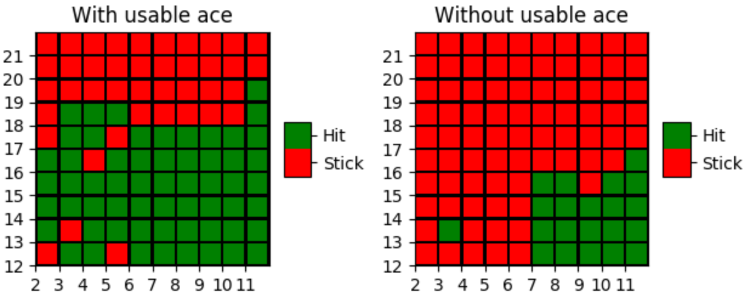


Figure 7: Policy after 500k episodes for $\alpha = 0.01$, $\gamma = 0.99$ ($\frac{1}{500000}$ chance of taking action with worse value)

5 QLearning

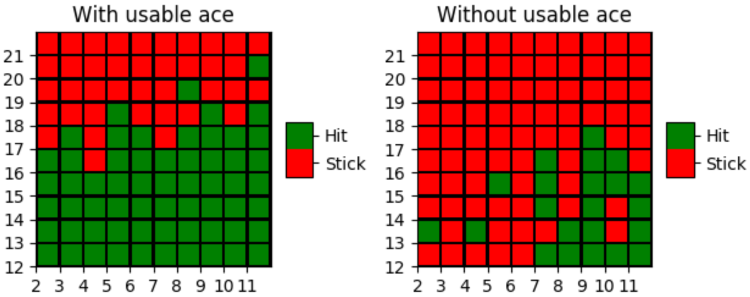
```
####################
QLearning (50000 epochs, 4.63s):
    GameStats(wins=0.4298, draws=0.0874, losses=0.4828) alpha: 0.01 gamma: 0.5
QLearning (50000 epochs, 4.43s):
    GameStats(wins=0.4239, draws=0.0886, losses=0.4875) alpha: 0.01 gamma: 0.9
QLearning (50000 epochs, 4.55s):
    GameStats(wins=0.4436, draws=0.0892, losses=0.4672) alpha: 0.01 gamma: 0.99
QLearning (50000 epochs, 4.48s):
    GameStats(wins=0.4162, draws=0.085, losses=0.4988) alpha: 0.1 gamma: 0.5
QLearning (50000 epochs, 4.51s):
    GameStats(wins=0.4216, draws=0.0913, losses=0.4871) alpha: 0.1 gamma: 0.9
QLearning (50000 epochs, 4.72s):
    GameStats(wins=0.4204, draws=0.0878, losses=0.4918) alpha: 0.1 gamma: 0.99
QLearning (50000 epochs, 4.81s):
    GameStats(wins=0.4168, draws=0.0822, losses=0.501) alpha: 0.5 gamma: 0.5
QLearning (50000 epochs, 4.81s):
    GameStats(wins=0.401, draws=0.0683, losses=0.5307) alpha: 0.5 gamma: 0.9
QLearning (50000 epochs, 4.79s):
    GameStats(wins=0.3975, draws=0.0808, losses=0.5217) alpha: 0.5 gamma: 0.99
QLearning (500000 epochs, 47.90s):
    GameStats(wins=0.4198, draws=0.0925, losses=0.4877) alpha: 0.01 gamma: 0.5
QLearning (500000 epochs, 48.65s):
    GameStats(wins=0.4169, draws=0.0927, losses=0.4904) alpha: 0.01 gamma: 0.9
QLearning (500000 epochs, 48.54s):
    GameStats(wins=0.4276, draws=0.096, losses=0.4764) alpha: 0.01 gamma: 0.99
QLearning (500000 epochs, 48.92s):
    GameStats(wins=0.4291, draws=0.0894, losses=0.4815) alpha: 0.1 gamma: 0.5
QLearning (500000 epochs, 48.18s):
    GameStats(wins=0.4099, draws=0.0875, losses=0.5026) alpha: 0.1 gamma: 0.9
QLearning (500000 epochs, 46.76s):
    GameStats(wins=0.4257, draws=0.0887, losses=0.4856) alpha: 0.1 gamma: 0.99
QLearning (500000 epochs, 47.45s):
    GameStats(wins=0.4046, draws=0.0776, losses=0.5178) alpha: 0.5 gamma: 0.5
QLearning (500000 epochs, 48.33s):
    GameStats(wins=0.4019, draws=0.0804, losses=0.5177) alpha: 0.5 gamma: 0.9
QLearning (500000 epochs, 48.20s):
    GameStats(wins=0.4039, draws=0.0835, losses=0.5126) alpha: 0.5 gamma: 0.99
```



With usable ace

Without usable ace

Figure 8: Policy after 50k episodes for $\alpha = 0.01,\ \gamma = 0.99$



With usable ace

Without usable ace

Figure 9: Policy after 500k episodes for $\alpha = 0.1,\ \gamma = 0.5$