Mission 2

Adam Morris – Computational Social Cognition Bootcamp, July 2017
thatadammorris.com thatadammorris@gmail.com

Rumors say the owners of the casino are not happy. Your RL agent has it too easy. In an attempt to befuddle your agent, the casino adds a second slot machine.

Congratulations on completing your first mission. If you want to compare your code to mine, my solution is in "slot_machine.m" (online here).

In Mission 2, you will modify your code to account for the second slot machine. On each round, your RL agent must now make two choices. First, it must choose which slot machine to go to. Then, it must choose which arm to pull at that machine. (Assume that the first machine has not changed from Mission 1, and that the second machine gives the following rewards: left arm = -2, right arm = +2.)

Part A

Simulate a Q-learning agent playing this game for 500 rounds, using the same learning rate and temperature as before. Make sure to do the proper updating of your Q values after each choice! Plot the results. (To make the plot this time, store your results in a 500 x 3 matrix, where each row is a different round, and the columns are: first choice, second choice, reward. Then call: `plotResults(results, 2)`.)

Try changing the reward of the left arm in the new slot machine from -2 to -20, or -200. Does this significantly affect the agent's behavior? Why or why not? What is the *specific* part of the code that allows agents to be robust to this manipulation? Do you think humans would be robust to this manipulation?

Part B

The casino has been subtly handing your agent cocktails all night, and now your agent is drunk. When it tries to navigate to a slot machine, it usually (70% of the time) gets there; but 30% of the time, its vision gets blurry and it accidentally stumbles over to the other slot machine. (Once it sits down at a slot machine, it knows which one it is playing. The stochasticity is just in the transitions.)

Implement this change in your code and rerun the simulations. (If you're having trouble, try drawing out the decision tree.) What is the best average reward that the agent can now achieve?