1. Consider the bar problem from Homework 3 with the $b=5$, $k=5$, and 50-agent case.

(a) Find the Nash Equilibrium when each agent uses the local reward:

$$L(z) = x_k(z)e^{-\frac{x_k(z)}{b}}$$

where $z$ is the system state and $x_z(z)$ is the number of agents that attend on night $k$. Then, find the Nash Equilibrium for the global reward, given by:

$$G(z) = \sum_{k=1}^{K} x_k(z)e^{-\frac{x_k(z)}{b}}$$

Are the Nash Equilibrium the same? What does the answer imply about the expected performance of the local reward?

(b) Now find the Nash Equilibrium for the Difference Reward with a zero counterfactual:

$$D_i(z) = x_i(z)e^{-\frac{x_i(z)}{b}} - (x_i(z) - 1)e^{-\frac{(x_i(z)-1)}{b}}$$

where $z$ is the system state, and $x_i(z)$ is the attendance of the night agent $i$ selected. Is this the same as the local or global reward’s Nash Equilibria? What does the answer imply about the expected performance of the local reward?

2. Now consider a new problem where two players are deciding which bar to attend. Player 1 prefers McMenamin’s, and Player 2 prefers Clod’s. However, neither player wants to drink alone. This is reflected in the payoff matrix, where the first entry is Player 1’s payoff:

<table>
<thead>
<tr>
<th></th>
<th>Clod’s</th>
<th>McMenamin’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clod’s</td>
<td>1,2</td>
<td>0,0</td>
</tr>
<tr>
<td>McMenamin’s</td>
<td>0,0</td>
<td>2,1</td>
</tr>
</tbody>
</table>

(a) Assume that Player 1 goes to McMenamin’s each time. Use a simple action-value learner to learn a policy for Player 2 that maximizes its payoff. How does this relate to the Nash Equilibrium of the problem?

(b) Now assume that Player 1 will choose McMenamin’s with probability $1/2$ and Clod’s with probability $1/2$. Learn a strategy to optimize Player 2’s payoff given that Player 1 has this mixed strategy. To what policy does your learner converge? Does that policy maximize Player 2’s payoff? Is this a Nash Equilibrium? Explain your answer.

(c) Now assume that Player 1 will choose McMenamin’s with probability $2/3$ and Clod’s with probability $1/3$. Learn a strategy to optimize Player 2’s payoff given that Player 1 has this mixed strategy. To what policy does your learner converge? Does that policy maximize Player 2’s payoff? Is this a Nash Equilibrium? Explain your answer.

(d) What is the mixed strategy equilibrium for this game? If one exists, how do the payoffs compare to the pure strategy equilibria payoffs?