CSE 473, Autumn 2018 – Project 3 **Name:**
Paul G. Allen School of Computer Science and Engineering

Project 3 - Question 9

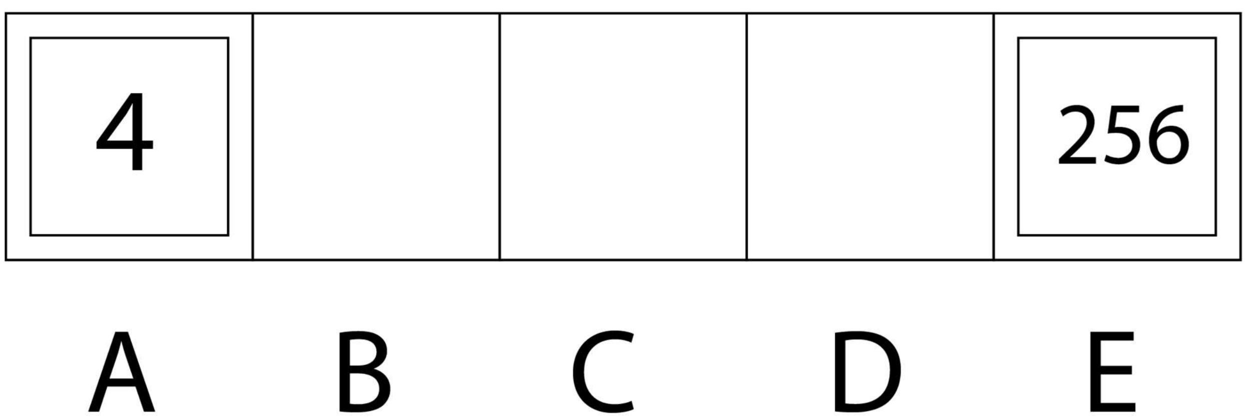
This non-programming problem is part of Project 3. Please add your answers to this document and submit your completed document along with your solution to the Pac-Man project.

The following problems take place in various scenarios of a 1D gridworld MDP.

In all cases double-rectangle states are exit states. From an exit state, the only action available is Exit, which results in the listed immediate reward and ends the game (by moving into a terminal state; not shown).

From non-exit states, the agent can choose either Left (L) or Right (R) actions, which move the agent in the corresponding direction. Assume that living rewards of -1 apply on transitions from states B, C, and D; the rewards from states A and E are as shown there, obtained by exiting the grid, which is the only permitted action in those states. Throughout this problem, assume that value iteration begins with initial values V0(s) = 0 for all states.

Consider the following scenario:



CSE 473, Autumn 2018 – Project 3 **Name:**
Paul G. Allen School of Computer Science and Engineering

Let the discount factor be γ = 0.5, and let transitions be deterministic. Fill in the missing values for each state following the value iteration algorithm in the following table:

Step t	$V_t(A)$	$V_t(B)$	$V_t(C)$	$V_t(D)$	$V_t(E)$
0	0	0	0	0	0
1					
2					
3					
4					

What are the actions for states B, C, and D for each time step, according to the best policy that corresponds to the values at step t?

Time	π_t (B)	π_t (C)	π_t (D)
1			
2			
3			
4			