Reinforcement Learning Example **Department of Computer Science**

Professor Carolina Ruiz **WPI**

Consider the deterministic grid world given below. Assume that:

- there are 4 available actions in a given state (= cell): up, down, left, right; except for actions that take the robot outside the grid world
- the immediate rewards are: 100 for actions that move the robot up or right; and 0 for actions that move the robot left or down
- the grid has two absorbing goal states G1 and G2, where the only available action is to stay in the same state with reward equal to 0
- the discount factor is $\gamma = 0.8$

Use the following grids to fill in the information requested on the top of the grid:

r(s,a): immediate reward value of action a

			G2
G1			

Q(s,a) value of each action

			G2
G1			

V*(s) value of each state

			G2
G1			

π*: An optimal policy

			G2
G1			

Now apply the Q learning algorithm to this grid world assuming that the table of Q-hat values is initialized to zero. Assume the robot begins in the bottom left corner and travels counterclockwise around the perimeter of the grid until it reaches an absorbing state, completing the 1st training episode. Describe what Q-hat values are modified and give their revised values. Perform a 2nd identical episode, and answer the same questions.

			G2
G1			