Solution Q2.

From the given fig, For each of the three states highlighted in green, we need to state what the policy action in that state is (e.g. UP, DOWN, LEFT, RIGHT).

For estimating such policy action we will be using **Bellman's Equation.** Its Eq is given as:-

$$\boxed{V(s) = \max\,[a]\{\mathrm{R}(\mathrm{s},\mathrm{a}) + \gamma * \sum \mathrm{P}(\mathrm{s}'|\mathrm{s},\mathrm{a}) * \mathrm{V}(\mathrm{s}')\}}$$

Where:

V(s) is the value function.

max[a] is the maximum operator,

R(s,a) is the immediate reward received for taking action a at state s.

γ (gamma) is the discount factor, a value between 0 and 1 that determines the importance of future rewards.

P(s'|s,a) is the probability of transitioning to state s' from state s when taking action a.

V(s') is the value function for the next state s'

Expected Utility: The expected utility formula is used to calculate the expected value of a decision in terms of utility. The expected utility formula is as follows:-

$$\boxed{\mathrm{EU}(\mathrm{a}) = \sum (\mathrm{u}(\mathrm{x}) * \mathrm{P}(\mathrm{x}\,|\mathrm{a}))}$$

Where:

EU(a) is the expected utility of action a

u(x) is the utility of outcome x

P(x|a) is the probability of outcome x given that action a is taken.

Optimal Policy Equation:

$$\boxed{\pi^*(s) = \mathrm{argmax}(a)\ Q^*(s,a)}$$

Where,
π*(s) is the optimal policy for a given state s.
argmax(a) represents the action that maximizes the value of Q*(s,a) .
Q*(s,a) is the optimal action-value function.

We need to calculate the policy action for each state in green colour, also given that if the agent attempts to move outside of the grid then it remains inside the grid in its position where it is.

1) We calculate the expected utility of (2,1)
   i) Expected Utility for Moving UP
$$Eu(up)_{(1,1)} = -0.1 + (7.63 * 0.8 + 7.63 * 0.1 + 7.40 * 0.1) = 7.507$$
   ii) Expected Utility of Moving Right
$$Eu(Right)_{(2,2)} = -0.1 + (7.63 * 0.8 + 7.63 * 0.1 + 6.30 * 0.1) = 7.397$$
   iii) Expected Utility of Moving Down
$$Eu(Down)_{(3,1)} = -0.1 + (6.30 * 0.8 + 7.63 * 0.1 + 7.40 * 0.1) = 6.443$$
   iv) Expected Utility of Moving Left
$$Eu(Left)_{(2,1)} = -0.1 + (7.40 * 0.8 + 7.63 * 0.1 + 6.30 * 0.1) = 7.213$$

The expected utility value of $Eu(up)_{(1,1)}$ higher i.e., 7.507.
Based optimal policy eq, we can conclude that the agent will move **UP**.

2) We calculate the expected utility of (1,4)
   i) Expected Utility of Moving UP
$$Eu(Up)_{(1,4)} = -0.1 + (8.05 * 0.8 + 8.18 * 0.1 + 7.93 * 0.1) = 7.951$$
   ii) Expected Utility of Moving Left
$$Eu(Left)_{(1,3)} = -0.1 + (7.93 * 0.8 + 8.05 * 0.1 + 8.05 * 0.1) = 7.854$$
   iii) Expected Utility of Moving Down
$$Eu(Down)_{(1,4)} = -0.1 + (8.05 * 0.8 + 7.93 * 0.1 + 8.18 * 0.1) = 7.951$$
   iv) Expected utility of Moving Right Block
$$Eu(Right)_{(1,5)} = -0.1 + (8.18 * 0.8 + 8.05 * 0.1 + 8.05 * 0.1) = 8.054$$

The expected utility value o$Eu(Right)_{(1,5)}$ higher i.e., 8.054.
Based optimal policy eq, we can conclude that the agent will move **RIGHT**

3) We calculate the expected utility of (4,3)
   i) Expected Utility of Moving UP
$$Eu(Up)_{(3,3)} = -0.1 + ((-3) * 0.8 + 6.06 * 0.1 + 7.90 * 0.1) = -1.104$$
   ii) Expected Utility of Moving Left
$$Eu(Left)_{(4,2)} = -0.1 + (6.06 * 0.8 + (-3) * 0.1 + 4.94 * 0.1) = 4.942$$
   iii) Expected Utility of Moving Down
$$Eu(Down)_{(5,3)} = -0.1 + (4.94 * 0.8 + 7.90 * 0.1 + 6.06 * 0.1) = 5.248$$

iv) Expected utility of Moving Right Block

$$Eu(Right)_{(4,4)} = -0.1 + (7.90 * 0.8 + (-3) * 0.1 + 4.94 * 0.1) = 6.414$$

The expected utility value o$Eu(Right)_{(4,4)}$ higher i.e., 6.414.
Based optimal policy eq, we can conclude that the agent will move **RIGHT**

C / R	1	2	3	4	5
1	7.63	7.78	7.93	8.05	8.18
2	7.40	7.63			8.32
3	6.30	6.34	-3	-3	8.44
4	5.96	6.06	4.94	7.90	10
5	5.82	5.92		3	9.11

Fig 1