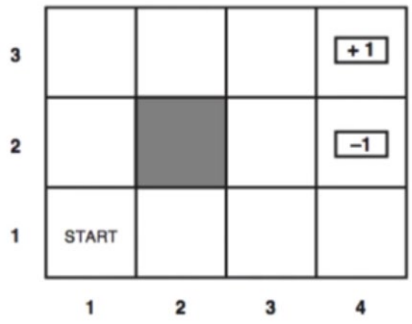
Teamwork assignment – solve the gridworld problem using Q-learning
Due on 10/1 midnight

Solve the 3 by 4 gridworld problem using Q-learning.
What to report:

- **Present your solution (using either an online or offline approach) along with a brief narrative. Include a link to your code so results can be verified.**
- **Discuss your findings on the first two questions below.**

1) How hyperparameters (learning rate α, discount factor γ, exploration schedule ε in case of online learning) affect learning?
2) Does Q value converge first or the policy converge first?
3) If you are curious, you may try to replace the penalty in state (4,2) from -1 by -200 and see how your solution is affected, and also if the solution makes sense when you compare the 2 cases (penalty -1 vs -200).

The gridworld:



- State transition: a desired direction occurs 80% of the time, 10% of the time to the left and 10% of the time to the right.
- Collision with walls results in staying at the same spot
- Two terminal states have reward +1 and -1, respectively.
- Each move at all other states, except the two terminal states, has a reward of -0.04
- The controls/actions at each state can be a move in one of the four directions N, S, W, E
- For example, in position (3,2), if the desired direction is N, then 80% of the time you will move in that direction, but 10% of the time you end up moving into the wall, while another 10% of the time you end up moving into the trap with a big penalty of -1.