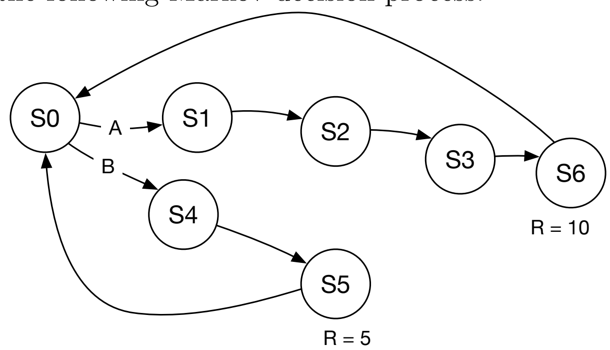
Name: ______________________

Murky decision problem

5. (8 points) Consider the following Markov decision process:



Assume:

- Reward is 0 in all states, except +10 in s6 and +5 in s5; the reward is received when *exiting* the state.
- Transitions out of s0 are deterministic, and depend on the choice of action (A or B).

(a) Assume in this part that all transitions are deterministic, following the arrows indicated with probability 1. When horizon = 3 and discount factor $\gamma = 1$, provide values for:

i. $Q(s_0, A)$ **0**

ii. $Q(s_0, B)$ **5**

(b) Still assuming that all transitions are deterministic, but letting horizon = 5 and discount factor $\gamma = 1$, provide values for:

i. $Q(s_0, A)$ **10**

ii. $Q(s_0, B)$ **5**

(c) Now, assume that transitions out of s_0 are deterministic, but that all other transitions follow the arrows indicated with probability 0.9 and stay in the current state with probability 0.1.

For policy $\pi(s_0) = B$, write a system of equations that can be solved in order to compute $V_\pi(s_0)$ when the horizon is infinite and $\gamma = 0.8$.

Do not solve the equations!

Solution:

$$v_0 = 0.8v_4$$
$$v_4 = 0.8(0.1v_4 + 0.9v_5)$$
$$v_5 = 5 + 0.8(0.1v_5 + 0.9v_0)$$