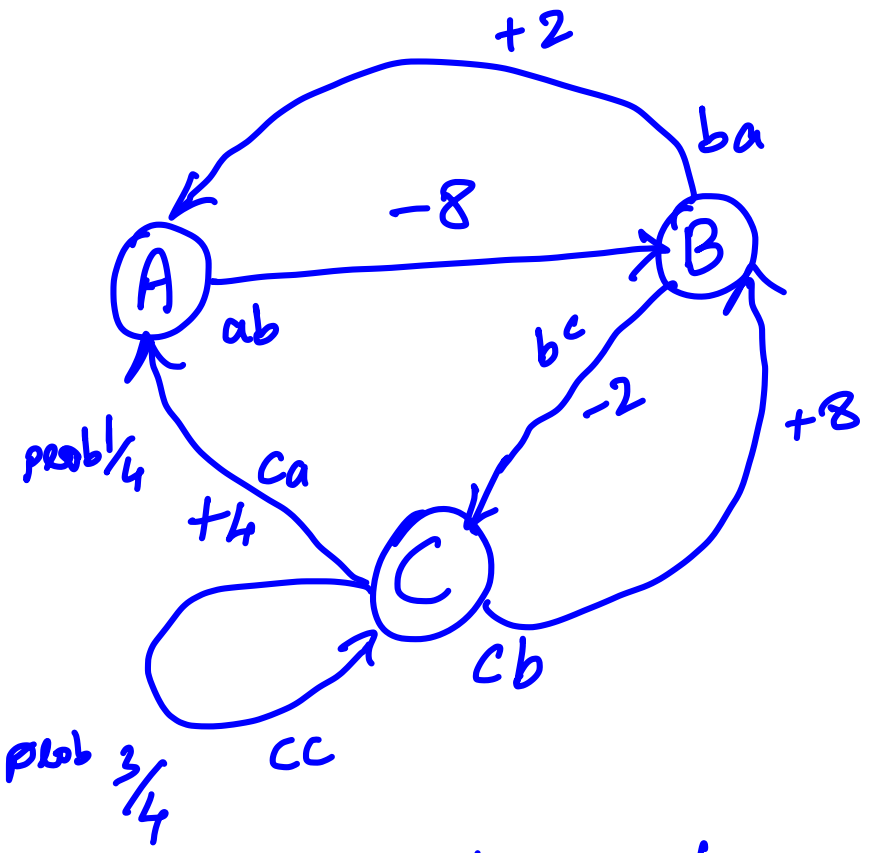
MDP $\gamma = 0.5$



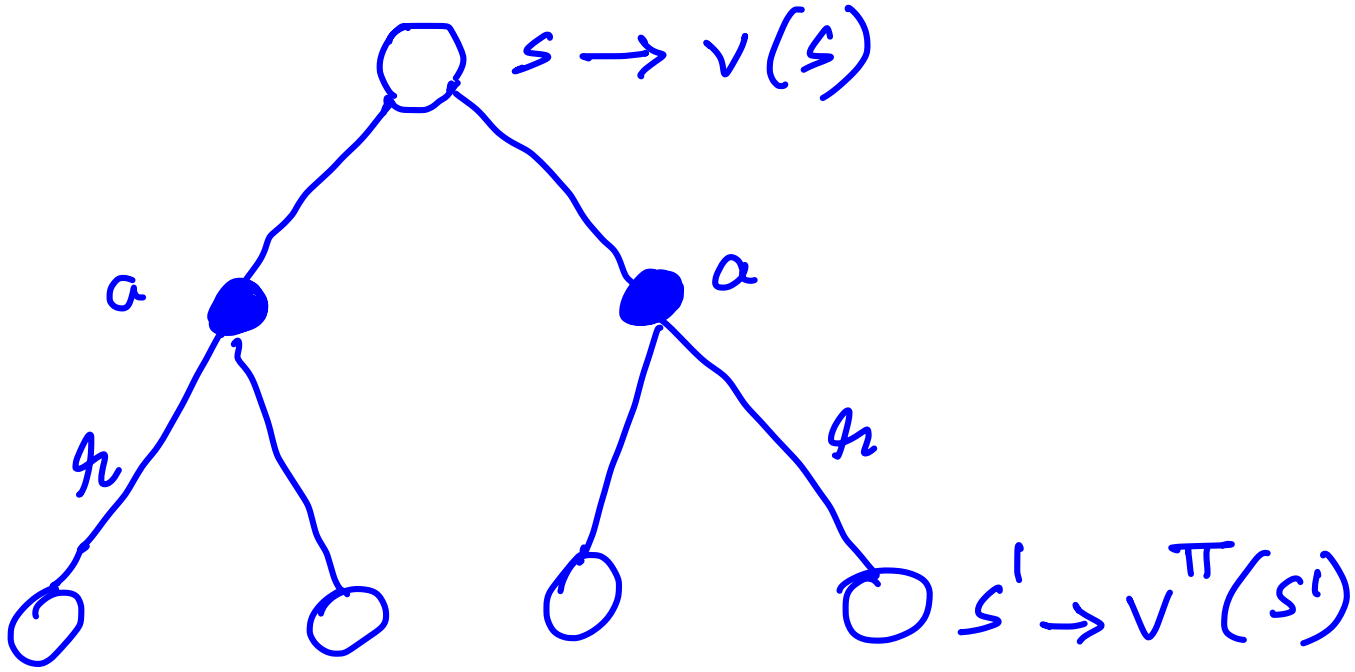
uniform random policy - all actions have equal probability

$V_1(A) = V_1(B) = V_1(C) = 2$

1 iteration of iterative policy evaluation (one back up for each state)

Compute new value function $V_2(s)$

$V_2(A)$, $V_2(B)$, $V_2(C) = ?$



$$V_k^{\pi}(s) = E\left[r + \gamma V^{\pi}(s')\right]$$

$$= \sum_{a \in A} \pi(a|s)\left(r + \gamma \sum_{s' \in S} P(s'|s,a) V_{k-1}^{\pi}(s')\right)$$

policy evaluation

i/p MDP $\langle S, A, P, R, \gamma \rangle$
π policy

o/p $V^{\pi}(s)$

actions → ab, ba, bc, ca, cb

$V_1(s)$

A	B	C
2	2	2

$V_2(s)$

A	B	C
-7	1	7

$$V_2(A) = -8 + 0.5(2) = -7$$

$$V_2(B) = 0.5(-2 + 0.5(2)) + 0.5(+2 + 0.5(2))$$

$$= -\frac{1}{2} + \frac{3}{2} = \frac{2}{2} = 1$$

$$V_2(B) = 0.5\left(8 + 0.5\left(2\right)\right)$$

$$+ 0.5\left(+4 + 0.5\left(0.25 * 2 + 0.75 * 2\right)\right)$$

$$= \frac{9}{2} + \frac{5}{2} = \frac{14}{2} = 7$$

$V_2(S)$