Sam Hopkins
CS 470
Hw 9 (MDP)

1 a) C $2 \to \{4(.8), 5(.2)\} = 100 \cdot .8 + 200 \cdot .2 = 80 + 40 = \underline{120}$
D $2 \to \{4(.6), 6(.4)\} = 100 \cdot .6 + 50 \cdot .4 = 60 + 20 = \underline{80}$

b)

1	2	3	4	5	6
0	0	0	100	200	50

for each
max (set of possibilities)

1	2	3	4	5	6
0	90	140	100	200	50

1: A = 0
B = 0

2: C = 90
D = 80

3: C = 50
D = 140

1	2	3	4	5	6
120	90	140	100	200	50

1: $A = .8 \cdot (90) + .2 \cdot (140) = 100$
$B = .4(90) + .6(140) = 120$

2: C = 90
D = 80

3: C = 50
D = 140

1	2	3	4	5	6
120	90	140	100	200	50

✓

c) optimal is $\pi^* = \{1(B), 3(D), 2(C)\}$

2 a) [-1 -1 0]

$s_1 \to a_1 = 1 \to s_1$
$a_2 = 1 \to .5(s_1), .5(s_2)$

$s_2 \to a_1 \to s_2 \to s_2$
$a_2 \to s_2 \to .5(s_3), .5(s_2)$

$s_3 \to s_3$

$\pi^*(s_1) = a_2$
$\pi^*(s_2) = a_2$
$\pi^*(s_3) = a_3$? (we only have one action)

b) $s_1 = a_1 + a_2 = -1 + (-1 \cdot .5 + -1.5 \cdot .5) = -2$ $\quad s_1 = R(s_1) + \Sigma 1$
$s_2 = a_1 + a_2 = -1 + (-1 \cdot .5 + 0 \cdot .5) = -1.5$
$s_3 = 0$

$V(1) = -4$
$V(2) = -2$
$V(3) = 0$

$V(3) = 0$
$V(2) = -1 + \frac{V(2)}{2} + \frac{V(3)}{2} = 2V(2) = -2 + V(2) + V(3)$
$\qquad - V(2)$
$V(2) = -2 + V(3)$
$\underline{V(2) = -2}$
$V(1) = -1 + \frac{V(1)}{2} + \frac{V(2)}{2}$
$2V(1) = -2 + V(1) + V(2)$
$\quad - V(1)$
$V(1) = -2 + V(2) = -2 + -2 = -4$

c) V^* would increase, as each iteration instead of multiplying by 1, we would multiply by .99. This makes $V(2)$ a bigger number (because it is negative, $-2 < -1.99$) which helps $V^*(1)$ increase as well.