Subject Abdelrahman Adel Abdelfattah Sheet 5 Date 17/12/96

① Policy Iteration

$A = \{B, O, D\}$, $S = \{N, F, A\}$

a) $|A|^{|S|} = 3^3 = 27$ Policies

b)

	N	F	A
π_0	O	B	D
V^{π_0}	$-\frac{1}{4}$	0	$-\frac{1}{100}$
~~π_1~~ π_1	B	B	B
V^{π_1}	0.113	-0.113	0.113
π_2	B	B	B

Done ✓

$V^{\pi_0}(NONE) = \frac{1}{4} - \frac{1}{2} + 0 = -\frac{1}{4}$

$V^{\pi_0}(F) = 0$

$V^{\pi_0}(A) =$

$\pi_1(None) = Max\{B: -0.113, O: -0.284, D: 0.131\}$

$\pi_1(F) = Max\{B: -0.113, O: -0.284, D: 0.131\}$

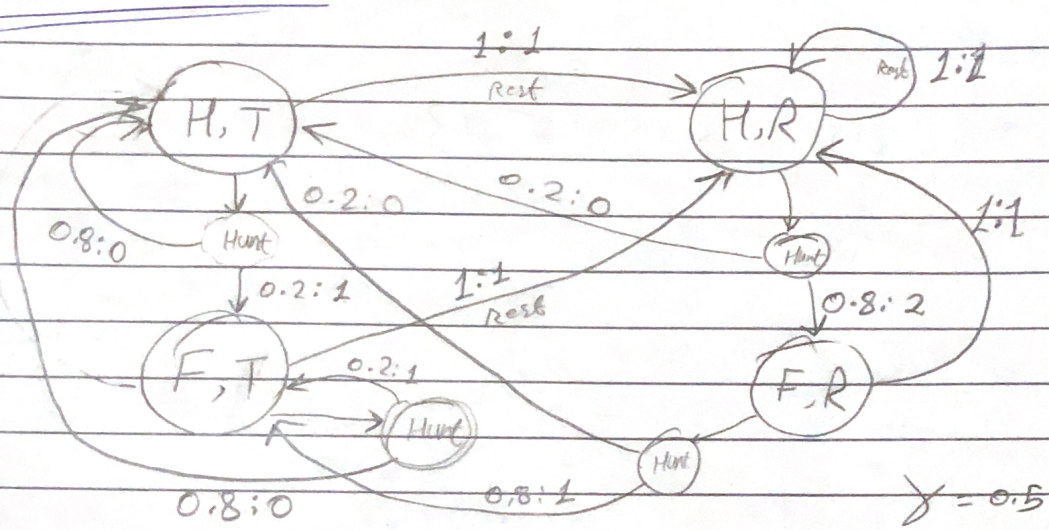
$\pi_1(A) = Max\{B: -0.113, O: -0.284, D: 0.131\}$

Best Policy is **BBB**

c) No, the Reward is a lot larger than the ~~γ~~ γV which will cause it to converge no matter the γ & also it's a cyclic graph all point to all

Subject _____ Date / /

② Value Iteration:-



a) $V^{\pi}(HUNgry) = \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^{\pi}(s')\right]$

$= 1 \times [1 + 0.5 \times V(H,R)]$

$= 1 + 0.5\, V(H,R)$

$V(H,R) = 1 \times [1 + 0.5\, V(H,R)]$

$= 1 + 0.5\, V(H,R)$

$V^{\pi}(H,T) = 1 + 0.5(1 + 0.5 V(H,R))$

$= 1 + 0.5 + 0.25 V(H,R)$

$= 1 + \gamma + \gamma^2 V(H,R)$

Geometric series sum

$$V^{\pi}(H,T) = \sum_{i=0} \gamma^i$$

$$V^{\pi}_{H,T} = \frac{1}{1-\gamma}$$ //

Subject _____ Date / /

b) $V_1(H,T) = \max\begin{cases}\text{Rest}: 1\\ \text{Hunt}: 0.2\end{cases}$

$= 1$

$V_1(H,R) = \max\begin{cases}\text{Rest}: 1\\ \text{Hunt}: 1.6\end{cases}$

$= 1.6$

$V_1(F,T) = \max\{\text{Rest}: 1, \text{Hunt}: 0.2\}$

$= 1$

$V_1(F,R) = \max\{\text{Rest}: 1, \text{Hunt}: 0.8\}$

$= 1$

	H, T	H, R	F, T	F, R
V_0	0	0	0	0
V_1	1	1.6	1	1
V_2	2.6	2.6	2.6	2.6
V_3	3.6	4.2	3.6	3.6
extra V_4	5.2	5.2	5.2	5.2

#