(1). a. $E[r^i] = p_i(1-p_i)$.

Arm 1 : $\frac{1}{4}(3/4)$

Arm 2 : $\frac{1}{3}(2/3)$. $\Rightarrow$ Arm 3 has highest expected reward.

Arm 3 : $1/2 \cdot 1/2$

Arm 4 : $2/3 \cdot 1/3$

Arm 5 : $3/4 \cdot 1/4$.

b. $\max_{p} \; p(1-p)$

$\Rightarrow \quad 1-2p = 0 \Rightarrow \underline{\underline{p = 1/2}}$

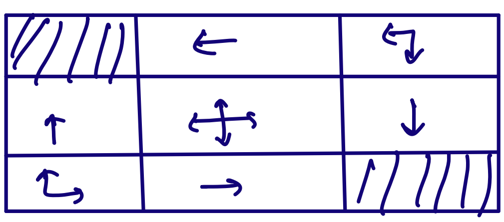
(2A). $V(s) = 0, \; \forall s = 1..7$. Take $\gamma = 1$ (any other γ is also acceptable).

step 1 :- $V(1) = \frac{3}{4}(-1) + \frac{1}{4}(-2) = -1.25 = V(7)$

$V(2) = \frac{2}{4}(-1) + \frac{2}{4}(-2) = -1.5 = V(6)$.

$V(3) = V(5) = \frac{3}{4}(-1) + \frac{1}{4}(-2) = -1.25$

$V(4) = -1$.



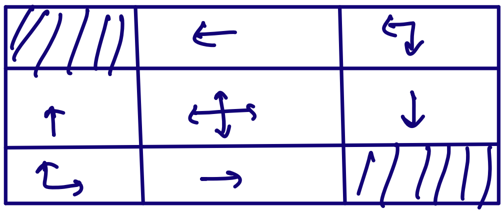
Step 2 :-

$V_2(1) = \frac{1}{4}(-1) + \frac{1}{4}(-2-1.25) + \frac{1}{4}(-1-1.5) + \frac{1}{4}(-1-1)$

$= -2.1875 = V_2(3) = V_2(5) = V_2(7)$

$$V_2(2) = \frac{2}{4}(-2-1.5) + \frac{2}{4}(-1-1.25) = -2.875 = V(6)$$

$$V_2(4) = -1-1.25 = -2.25$$



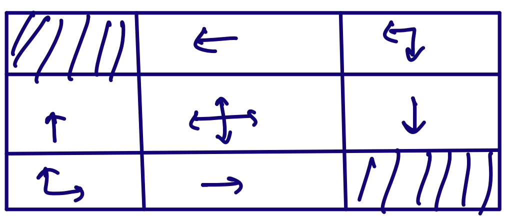
Step 3 :-

$$V_3(1) = \frac{1}{4}(-1) + \frac{1}{4}(-2-2.1875) + \frac{1}{4}(-1-2.875) + \frac{1}{4}(-1-2.25) = -3.078 = V_3(7) = V_3(5) = V_3(3).$$

$$V_3(2) = \frac{2}{4}(-2-2.875) + \frac{2}{4}(-1-2.1875)$$

$$= -4.03125 = V_3(6)$$

$$V_3(4) = -3.1875$$



(3A) (a). $$v^*(high) = \max_{\{search,\, wait\}} \left\{ \alpha[r_{search} + \gamma v^*(high)] + (1-\alpha)[r_{search} + \gamma v^*(low)],\ r_{wait} + \gamma v^*(high) \right\}.$$

$$v^*(low) = \max_{\{recharge,\, search,\, wait\}} \left\{ 0 + \gamma v^*(high),\ \beta[r_{search} + \gamma v^*(low)] + (1-\beta)[-3 + \gamma v^*(high)], \right.$$

$r_{wait} + \gamma v^{*}(low). \}.$

(b). $v^{\pi}(low) = \gamma v^{\pi}(high). = 0.9 v^{\pi}(high).$

$$v^{\pi}(high) = 5 + \gamma [0.3 v^{\pi}(high) + 0.7 v^{\pi}(low)]$$

$$= 5 + 0.27 v^{\pi}(high) + 0.63 v^{\pi}(low).$$

$$= 5 + 0.27 v^{\pi}(high) + 0.567 v^{\pi}(high).$$

$$= 5 + 0.837 v^{\pi}(high).$$

$$\Rightarrow \quad v^{\pi}(high) = \frac{5}{0.163} \simeq 30.67.$$

$$v^{\pi}(low) \simeq 27.61.$$