$S = \{\text{Sleep}, \text{Active}\}$, $A = \{\text{turn-ON}, \text{turn-OFF}, \text{Stay-ON}, \text{stay-OFF}\}$

$R = \{-1, 0, 1\}$

- $P(s' = \text{sleep} \mid s = \text{Active}, a = \text{turn-OFF}) = 1$
  - $\rightarrow P(r = 0 \mid s = \text{Active}, a = \text{turn-OFF}, s' = \text{Sleep}) = 0.3$
  - $\rightarrow P(r = -1 \mid s = \text{Active}, a = \text{turn-OFF}, s' = \text{Sleep}) = 0.7$
  - $\rightarrow P(r = 1 \mid S = \text{Active}, a = \text{turn-OFF}, s' = \text{Sleep}) = 0$

- $P(s' = \text{Active} \mid s = \text{Active}, a = \text{stay-ON}) = 0.6$
  - $\rightarrow P(r = 1 \mid S = \text{Active}, a = \text{stay-ON}, s' = \text{Active}) = 0.5$
  - $\rightarrow P(r = 0 \mid S = \text{Active}, a = \text{Stay-ON}, s' = \text{Active}) = 0.5$
  - $\rightarrow P(r = -1 \mid S = \text{Active}, a = \text{Stay-ON}, s' = \text{Active}) = 0$

- $P(s' = \text{sleep} \mid S = \text{Active}, a = \text{stay-ON}) = 0.4$
  - $\rightarrow P(r = 0 \mid s = \text{Active}, a = \text{stay-ON}, s' = \text{sleep}) = 0.6$
  - $\rightarrow P(r = -1 \mid s = \text{Active}, a = \text{stay-ON}, s' = \text{Sleep}) = 0.4$
  - $\rightarrow P(r = 1 \mid S = \text{Active}, a = \text{stay-ON}, s' = \text{Sleep}) = 0$

- $P(s' = \text{Sleep} \mid S = \text{Sleep}, a = \text{stay-OFF}) = 1$
  - $\rightarrow P(r = 0 \mid S = \text{Sleep}, a = \text{stay-OFF}, s' = \text{Sleep}) = 0.5$
  - $\rightarrow P(r = -1 \mid S = \text{Sleep}, a = \text{stay-OFF}, s' = \text{Sleep}) = 0.5$
  - $\rightarrow P(r = 1 \mid S = \text{sleep}, a = \text{stay-OFF}, s' = \text{sleep}) = 0$

- $P(s' = \text{Active} \mid S = \text{Sleep}, a = \text{turn-ON}) = 1$
  - $\rightarrow P(r = 0 \mid s = \text{Sleep}, a = \text{turn-ON}, s' = \text{Active}) = 0.8$
  - $P(r = -1 \mid S = \text{Sleep}, a = \text{turn-ON}, s' = \text{Active}) = 0.1$
  - $P(r = 1 \mid S = \text{Sleep}, a = \text{turnON}, s' = \text{Active}) = 0.1$

$\rightarrow P(r \mid S = \text{Sleep}, a, s') = 0 \quad \forall r, s'$
$S = \text{Sleep}, a = \{\text{turn-OFF}, \text{Stay-ON}\}$

- $P(s' \mid S = \text{Sleep}, a = \{\text{turn-OFF}, \text{stay-ON}\}) = 0$

$P(s' \mid S = \text{Active}, a = \{\text{turn-ON}, \text{Stay-OFF}\}) = 0 \longrightarrow P(r \mid s, a, s') = 0 \quad \forall r, \forall s'$
$S = \text{Active}$
$a = \{\text{turn-ON}, \text{Stay-OFF}\}$

$$- P(s' \mid s = \text{Sleep}, a \in A') = 0 \rightarrow P(r \mid s = \text{sleep}, a \in A', s') = 0$$

$$\forall r \in R, \forall s' \in S, \; A' = \{\text{turn-OFF}, \text{Stay-ON}\}$$

$$- P(s' \mid s = \text{Active}, a \in A') = 0 \rightarrow P(r \mid s = \text{Active}, a \in A', s') = 0$$

$$\forall r \in R, \forall s' \in S, \; A' = \{\text{turn-ON}, \text{Stay-OFF}\}$$

(a)

$$P(s', r \mid s, a) = \frac{P(r, s, a, s')}{P(s,a)} = \frac{P(r \mid s, a, s') P(s, a, s')}{P(s,a)} =$$

$$\frac{P(r \mid s, a, s') \, P(s' \mid s, a) \, \cancel{P(s,a)}}{\cancel{P(s,a)}}$$

$$= P(r \mid s, a, s') \, P(s' \mid s, a)$$

(b) $$P(r \mid s, a) = \sum_{s' \in S} P(r, s' \mid s, a)$$

(c) $$P(s' \mid s, a) = \sum_{r \in R} P(r, s' \mid s, a)$$

(d) $$Q(s,a) = \mathbb{E}[r + V(s')] = \mathbb{E}[r] + \mathbb{E}[V(s')] = \sum_{r \in R} r \, P(r \mid s, a) + \mathbb{E}[V(s')]$$

$$= \sum_{r \in R} \sum_{s' \in S} r \, P(r, s' \mid s, a) + \sum_{s' \in S} V(s') \, P(s' \mid s, a)$$

$$= \sum_{r \in R} \sum_{s' \in S} r \, P(r, s' \mid s, a) + \sum_{r \in R} \sum_{s' \in S} V(s') \, P(r, s' \mid s, a)$$

$$= \sum_{r \in R} \sum_{s' \in S} (r + V(s')) \, P(r, s' \mid s, a)$$

(e) $V(s) = \max_a Q(s,a) = \max_a \sum_{r \in R} \sum_{s' \in S} (r + V(s'))\, p(r, s' \mid s, a)$

- $V(\text{Sleep}) = \max_a Q(s = \text{Sleep}, a) = \max_a \sum_{r \in R} \sum_{s' \in S} (r + V(s'))\, p(r, s' \mid s = \text{sleep}, a)$

  - Since there are only two possible actions to take at state $s = \text{Sleep}$, therefore

- $V(\text{Sleep}) = \max\left(Q(s = \text{Sleep}, a = \text{turn-ON}),\ Q(s = \text{Sleep}, a = \text{stay-OFF}) \right)$

$V(\text{Sleep}) = \max\left(V(\text{Active}),\ -0.5 + V(\text{Sleep}),\ 0 \right)$

$V(\text{Active}) = \max_a Q(s = \text{Active}, a) = \max_a \sum_{r \in R} \sum_{s' \in S} (r + V(s'))\, p(r, s' \mid s = \text{Active}, a)$

- Again, there are only two possible actions to take in state $S = \text{Active}$, so

  - $V(\text{Active}) = \max\left(Q(S = \text{Active}, a = \text{turn-OFF}),\ Q(s = \text{Active}, a = \text{stay-ON}) \right)$

$V(\text{Active}) = \max\left(-0.7 + V(\text{Sleep}),\ 0.14 + 0.4\, V(\text{Sleep}) + 0.6\, V(\text{Active}),\ 0 \right)$

(f) $$P_\mu(s'|s) = \sum_{a\in A} P(s'|s,a)\, \underbrace{P_\mu(a|s)}_{\mu(a|s)}$$

(g) • $P(s'=\text{Active} \mid s=\text{Sleep}) = 0.5$

$$= \sum_{a\in A} P(s'=\text{Active} \mid s=\text{Sleep}, a)\, \mu(a \mid s=\text{Sleep})$$

$$= \overset{1}{P(s'=\text{Active} \mid s=\text{Sleep}, a=\text{turn-ON})}\, \mu(a=\text{turn-ON} \mid s=\text{Sleep}) +$$
$$\overset{0}{P(s'=\text{Active} \mid s=\text{Sleep}, a=\text{stay-OFF})}\, \mu(a=\text{stay-OFF} \mid s=\text{Sleep})$$

$$= \mu(a=\text{turn-ON} \mid s=\text{Sleep}) = \frac{1}{2}$$

• $P(s'=\text{Sleep} \mid s=\text{Active}) = 0.5$

$$= \sum_{a\in A} P(s'=\text{sleep} \mid s=\text{Active}, a)\, \mu(a \mid s=\text{Active})$$

$$= \overset{1}{P(s'=\text{Sleep} \mid S=\text{Active}, a=\text{turn-OFF})}\, \mu(a=\text{turn-OFF} \mid s=\text{Active}) +$$
$$\overset{0.4}{P(s'=\text{Sleep} \mid s=\text{Active}, a=\text{stay-ON})}\, \mu(a=\text{stay-ON} \mid s=\text{Active})$$

$$= \frac{4}{10}\, \mu(a=\text{stay-ON} \mid S=\text{Active}) + \mu(a=\text{turn-OFF} \mid S=\text{Active}) = 1/2$$

★ one possible solution is $\begin{cases} \mu(a=\text{Stay-ON} \mid S=\text{Active}) = 0 \\ \mu(a=\text{turn-OFF} \mid s=\text{Active}) = 1/2 \end{cases}$

Policy:
$\mu(a=\text{turn-ON} \mid s=\text{Sleep}) = 0.5$
$\mu(a=\text{stay-OFF} \mid S=\text{Sleep}) = x, \quad x \in [0,1]$
$\mu(a=\text{turn-OFF} \mid s=\text{Active}) = 0.5$
$\mu(a=\text{stay-ON} \mid S=\text{Active}) = 0$