The MDP has 2 policies:

Policy 1:

S_1	a
S_2	d
S_3	d
S_4	d
S_5	d

Policy 2:

S_1	b
S_2	d
S_3	d
S_4	d
S_5	d

(XX) At infinite time, the value of $V(S_i)$ will get constant.

So, we can write the following equations:
for action a [when S1 is following policy 1]

$$V(S_1) = 0 + \gamma\left[(0.1)V(S_1) + (0.9)\cdot V(S_2)\right]$$

$$V(S_2) = 0 + \gamma\left[P_S \cdot V(S_2) + (1-P_S)V(S_4)\right]$$

$$V(S_3) = 0 + \gamma\left[(0.1)V(S_3) + P_R \cdot V(S_5) + (0.9-P_R)\cdot V(S_4)\right]$$

$$V(S_4) = 10 + \gamma\left[(0.1)\cdot V(S_4) + 0.9\cdot V(S_1)\right]$$

$$V(S_5) = -10 + \gamma\left[(0.1)\cdot V(S_5) + 0.9\cdot V(S_1)\right]$$

Solving them,

→ $V(S_2) = \gamma P_S \cdot V(S_2) + \gamma(1-P_S)V(S_4)$

$$\therefore V(S_2) = \frac{\gamma(1-P_S)V(S_4)}{(1-\gamma P_S)} \quad \text{— (1)}$$

→ $V(S_4) = 10 + \gamma(0.1)V(S_4) + \gamma\cdot(0.9)V(S_1)$

$$\therefore V(S_4) = \frac{\left[10 + (0.9)\gamma\cdot V(S_1)\right]}{1-(0.1)\gamma} \quad \text{— (2)}$$

Placing the value of $V(S_4)$ from equation (2) in equation (1), we get

$$V(S_2) = \frac{(1-P_S)\cdot\gamma\cdot(10+0.9\gamma\cdot V(S_1))}{(1-\gamma P_S)\cdot(1-0.1\gamma)} \quad \text{——(3)}$$

Placing this value in equation of $V(S_1)$, we get

$$V(S_1) = 0 + \gamma[(0.1)V(S_1)+(0.9)\cdot V(S_2)]$$

$$= \gamma(0.1)\cdot V(S_1) + \frac{\gamma^2(0.9)\cdot(1-P_S)(10+0.9\gamma\cdot V(S_1))}{(1-\gamma P_S)(1-0.1\gamma)}$$

$$\therefore (1-(0.1)\gamma)^2 V(S_1)\cdot(1-\gamma P_S) = \gamma(0.9-0.9P_S)(10+0.9\gamma\cdot V(S_1))$$

$$\therefore V(S_{10}) = \frac{(9-9P_S)\gamma^2}{(1-0.1\gamma)^2(1-\gamma P_S) - 0.81(\gamma^3-P_S\gamma^3)} \quad \text{——(4)}$$

We can similarly write the following equations for action b [i.e. when S_1 is following policy 2]

$$V(S_1) = 0 + \gamma[(0.1)V(S_1) + (0.9)V(S_3)]$$

$$V(S_2) = 0 + \gamma[P_S V(S_2) + (1-P_S)V(S_4)]$$

$$V(S_3) = 0 + \gamma[(0.1)V(S_3) + P_R \cdot V(S_5) + (0.9 - P_R)V(S_4)]$$

$$V(S_4) = 10 + \gamma[(0.1)V(S_4) + (0.9)V(S_1)]$$

$$V(S_5) = -10 + \gamma[(0.1)V(S_5) + (0.9)V(S_1)]$$

Solving them,

$$\rightarrow V(S_5) = -10 + \gamma(0.1)\cdot V(S_5) + \gamma(0.9)\cdot V(S_1)$$

$$\therefore V(S_5) = \frac{-10 + \gamma(0.9)\cdot V(S_1)}{1-(0.1)\gamma} \quad \text{——— (5)}$$

$$\rightarrow V(S_3) = \gamma(0.1)\cdot V(S_3) + \gamma P_R V(S_5) + \gamma(0.9-P_R)V(S_4)$$

Putting the values of $V(S_4)$ and $V(S_5)$ from equations (2) and (5), we get

$$V(S_3) = \gamma \cdot (0.1) V(S_3) + \gamma P_R \cdot \left(\frac{-10 + 0.9\gamma V(S_1)}{1 - 0.1\gamma} \right)$$

$$+ \gamma (0.9 - P_R) \left(\frac{10 + 0.9\gamma \cdot V(S_1)}{1 - 0.1\gamma} \right)$$

$$\therefore V(S_3) = \frac{\left[\gamma P_R (-10 + 0.9\gamma V(S_1)) + \gamma (0.9 - P_R)(10 + 0.9\gamma V(S_1)) \right]}{(1 - 0.1\gamma)^2} \quad \text{——— (6)}$$

$$\rightarrow \quad V(S_1) = \gamma (0.1) V(S_1) + \gamma (0.9) \cdot V(S_3)$$

Placing the value of $V(S_3)$ from equation (6), we get

$$(1 - 0.1\gamma) S_1 = \gamma \cdot 0.9 \left[\frac{\gamma P_R (-10 + 0.9\gamma V(S_1) + \gamma (0.9 - P_R)(10 + 0.9\gamma V(S_1))}{(1 - 0.1\gamma)^2} \right]$$

$$\therefore V(S_1) = \frac{\gamma (0.9) \left[-20\gamma P_R + 9\gamma + 0.81\gamma^2 V(S_1) \right]}{(1 - 0.1\gamma)^3}$$

$$V(S_1).(1-0.1\gamma)^3 - \gamma(0.9)(0.81\gamma^2 V(S_1)) = \gamma(0.9)(-20\gamma P_R + 9\gamma)$$

$$\therefore V(S_{1b}) = \frac{0.9\gamma^2(9-20P_R)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3} \quad \text{---------- (7)}$$

(1) (a) $P_S = 0.2$, $P_R = 0.01$

From eq.(4), we get $V(S_1)$ when following the first policy.

$$V(S_{1a}) = \frac{(9-9P_S)\gamma^2}{(1-0.1\gamma)^2((1-P_S\cdot\gamma) - 0.81(\gamma^3 - P_S\gamma^3)}$$

$$= \frac{6.498}{0.10783} = 60.26082242$$

Similary, if S_1 is following ~~policy~~ second policy, from eq.(7)

$$V(S_{1b}) = \frac{0.9\gamma^2(9-20P_R)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$= \frac{7.1478}{0.11619125} = 61.51754111$$

As $V(S_{1a}) < V(S_{1b})$, we select the second policy.

$\therefore V(S_1) = V(S_{1b}) = 61.517541 1$

From eq (2)

$$V(S_4) = \frac{10 + 0.9\gamma * V(S_1)}{1 - (0.1)\gamma}$$

$$= 69.168505 6801$$

From eq (5)

$$V(S_5) = \frac{-10 + \gamma(0.9) \cdot V(S_1)}{1 - (0.1)\gamma}$$

$$= 47.0690581663$$

From eq. (1)

$$V(S_2) = \frac{\gamma(1-P_S) \cdot V(S_4)}{(1 - \gamma P_S)}$$

$$= 64.8988448356$$

$$V(S_3) = \gamma[(0.1) V(S_3) + P_R \cdot V(S_5) + (0.9 - P_R) \cdot V(S_4)]$$

$$= 65.1150581272$$

∴ For $P_S = 0.2$, $P_R = 0.01$,

The optimal policy is

S_1	b
S_2	d
S_3	d
S_4	d
S_5	d.

and value function is

$V(S_1) = 61.51754111$
$V(S_2) = 64.89884448356$
$V(S_3) = 65.11505812 72.$
$V(S_4) = 69.16850568 01$
$V(S_5) = 47.06905 8166.$

(b) $P_S = 0.2$ $P_R = 0.03$.

$V(S_{1a}) = 60.26082442$ (from eq.(4))

from eq.(7),

$$V(S_{1b}) = \frac{0.9\gamma^2(9 - 20(0.03)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$= \frac{6.8229}{0.11619125} = 58.721289 3$$

As $V(S_{1a}) > V(S_{1b})$ we select first policy

$\therefore V(S_1) = V(S_{1a}) = 60.26082422$

From eq. (2),

$V(S_4) = 67.98122$

From eq (5),

$V(S_5) = 45.88177314$

From eq. (1)

$V(S_2) = 63.78484901$

$V(S_3) = 63.52940266$

$\therefore$ For $P_S = 0.2$ & $P_R = 0.03$,

optimal policy :

S_1	a
S_2	d
S_3	d
S_4	d
S_5	d

Value function is:

$V(S_1) = 60.26082422$

$V(S_2) = 63.78484901$

$V(S_3) = 63.52940266$

$V(S_4) = 67.98122$

$V(S_5) = 45.88177314$

(C) If the agent is indifferent between taking action a and b in state S_1, the ~~the~~ value function of S_1 should remain the same.

$$\therefore V(S_{1a}) = V(S_{1b})$$

$$\therefore \frac{(9 - 9P_S)\gamma^2}{(1-0.1\gamma)^2(1-\gamma P_S) - 0.81(\gamma^3 - P_S\gamma^3)}$$

$$= \frac{0.9\gamma^2(9-20P_R)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$\therefore 60.26082442 = \frac{(9-20P_R)\cdot 0.9\gamma^2}{0.11619125}$$

$$\therefore (9 - 20P_R) = 8.62022837$$

$$\therefore P_R = 0.018988585815$$

(2)(a) $P_S = 0.6 \quad P_R = 0.1$

from eq.(4), we get $V(S_1)$ when following the first policy as follows

$$V(S_{1a}) = \frac{(4 - 4P_S)\gamma^2}{(1-0.1\gamma)^2(1-P_S\gamma) - 0.81(\gamma^3 - P_S\gamma^3)}$$

$$= \frac{3.249}{0.35218075 - 0.2777895}$$

$$= 43.6744913.$$

Similarly, if S_1 is following second policy, we get the following from eq.(7)

$$V(S_{1b}) = \frac{0.9\gamma^2(9 - 20P_R)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$= 48.9344077.$$

$\therefore V(S_{1b}) > V(S_{1a})$

We select the second policy

$V(S_1) = V(S_{1b}) = 48.9344077.$

$$V(S_4) = \frac{10 + 0.9\gamma * V(S_1)}{1-(0.1)\gamma}$$

$$= 57.28057303.$$

$$V(S_5) = \frac{-10 + \gamma(0.9)V(S_1)}{1-(0.1)\gamma}$$

$$= 35.18112552.$$

$$V(S_2) = \frac{\gamma(1-P_S)V(S_4)}{(1-\gamma P_S)}$$

$$= 50.62004128$$

$$V(S_3) = \gamma[(0.1)V(S_3) + P_R V(S_5) + (0.9-P_R)V(S_4)]$$

$$= 51.79606898.$$

∴ for $P_S = 0.6$ and $P_R = 0.1$,

the optimal policy is

S_1	b
S_2	d
S_3	d
S_4	d
S_5	d

And the value function is

$V(S_1) = 48.93440747$
$V(S_2) = 50.62004128$
$V(S_3) = 51.79606898$
$V(S_4) = 57.28057303$
$V(S_5) = 35.18112552$

(b) $P_S = 0.6 \quad P_\gamma = 0.2$

From eq. (4)

$V(S_{1a}) = 43.6744913$

From eq. (7),

$$V(S_{1b}) = \frac{0.9\gamma^2(q - 20(0.2))}{(1 - 0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$= \frac{4.06125}{0.741217625 - 0.62502637 5}$$

$$= 34.95 31484$$

As $V(S_{1a}) > V(S_{1b})$, we select the first policy.

$\therefore V(S_1) = V(S_{1a}) = 43.6744913$

From eq(2),

$$V(S_4) = \frac{10 + 0.9\gamma \cdot V(S_1)}{1-(0.1)\gamma}$$

$$= 52.311259 72$$

From eq. (5),

$$V(S_5) = \frac{-10 + \gamma(0.9) \cdot V(S_1)}{1 - 0.1\gamma}$$

$$= 30.2118122 1.$$

From eq. (1)

$$V(S_2) = \frac{\gamma(1-P_S)V(S_4)}{(1-\gamma P_S)}$$

$$= 46.228555 11$$

$$V(S_3) = \gamma[(0.1)V(S_3) + P_R V(S_5) + (0.9-P_R)V(S_4)]$$

$$= 44.781477186$$

$\therefore$ f

∴ for $P_S = 0.6$, $P_R = 0.2$,

the optimum policy is

S_1 a
S_2 d
S_3 d
S_4 d
S_5 d

And the value function is:

$V(S_1) = 43.67449129$
$V(S_2) = 46.22855511$
$V(S_3) = 42.78147186$
$V(S_4) = 52.31125972$
$V(S_5) = 30.21181221$

(C) If the agent is indifferent, then

$$V(S_{1a}) = V(S_{1b})$$

$$\therefore \frac{(9 - 9P_S)\gamma^2}{(1-0.1\gamma)^2(1-\gamma P_S) - 0.81(\gamma^3 - P_S\gamma^3)}$$

$$= \frac{0.9\gamma^2(9 - 20P_R)}{(1-0.1\gamma)^3 - (0.9)^3\gamma^3}$$

$$\therefore 43.67449913 = \frac{(0.9)(0.95)^2(9 - 20P_R)}{0.11619125}$$

$$\therefore P_R = 0.137621192$$