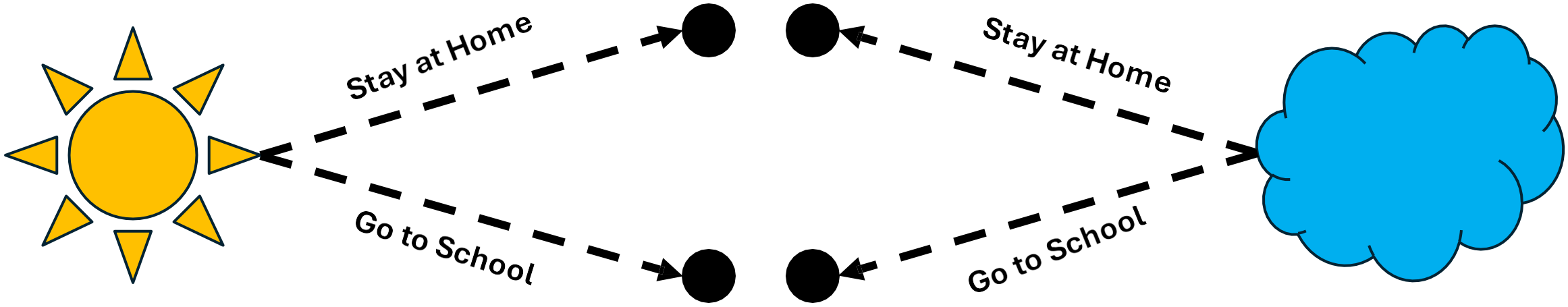
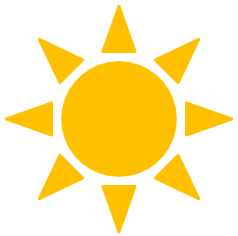
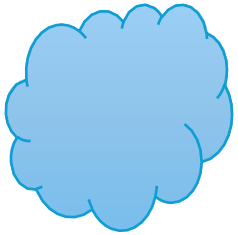
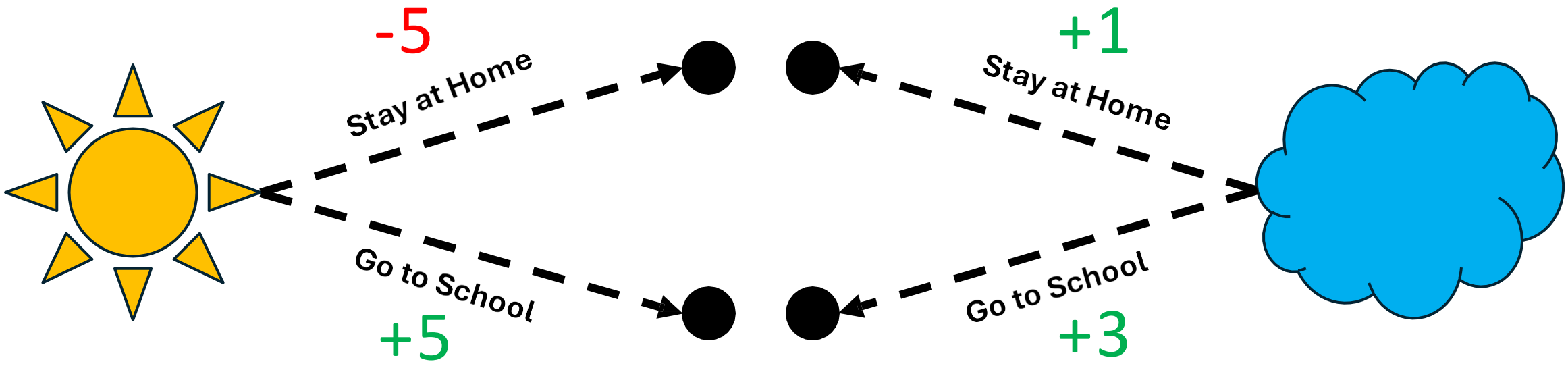


States s: {Sunny, Cloudy}

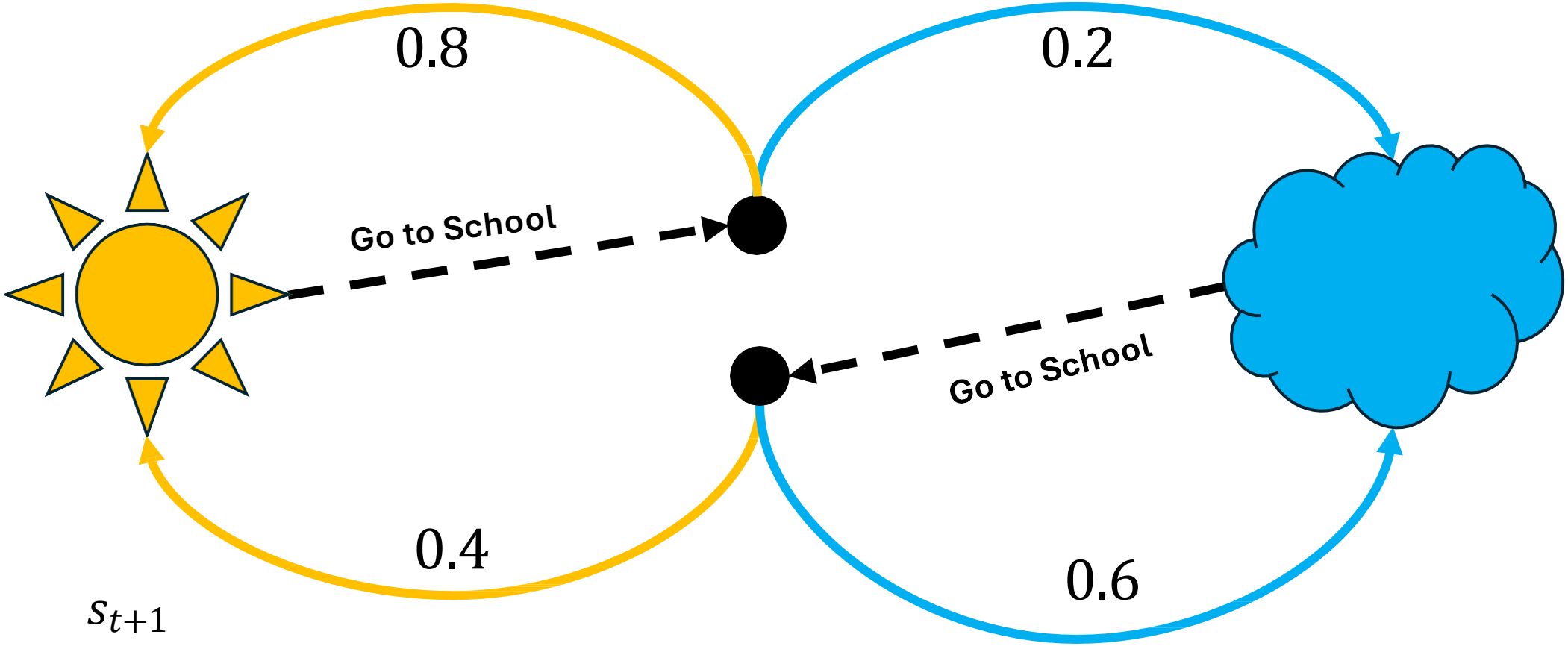
Actions a: {Go to School, Stay at Home}

Discount $\gamma = 0.9$

	Go to School	Stay at Home
	+5	-5
	+3	+1



$$R_{school} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \quad R_{stay} = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$$

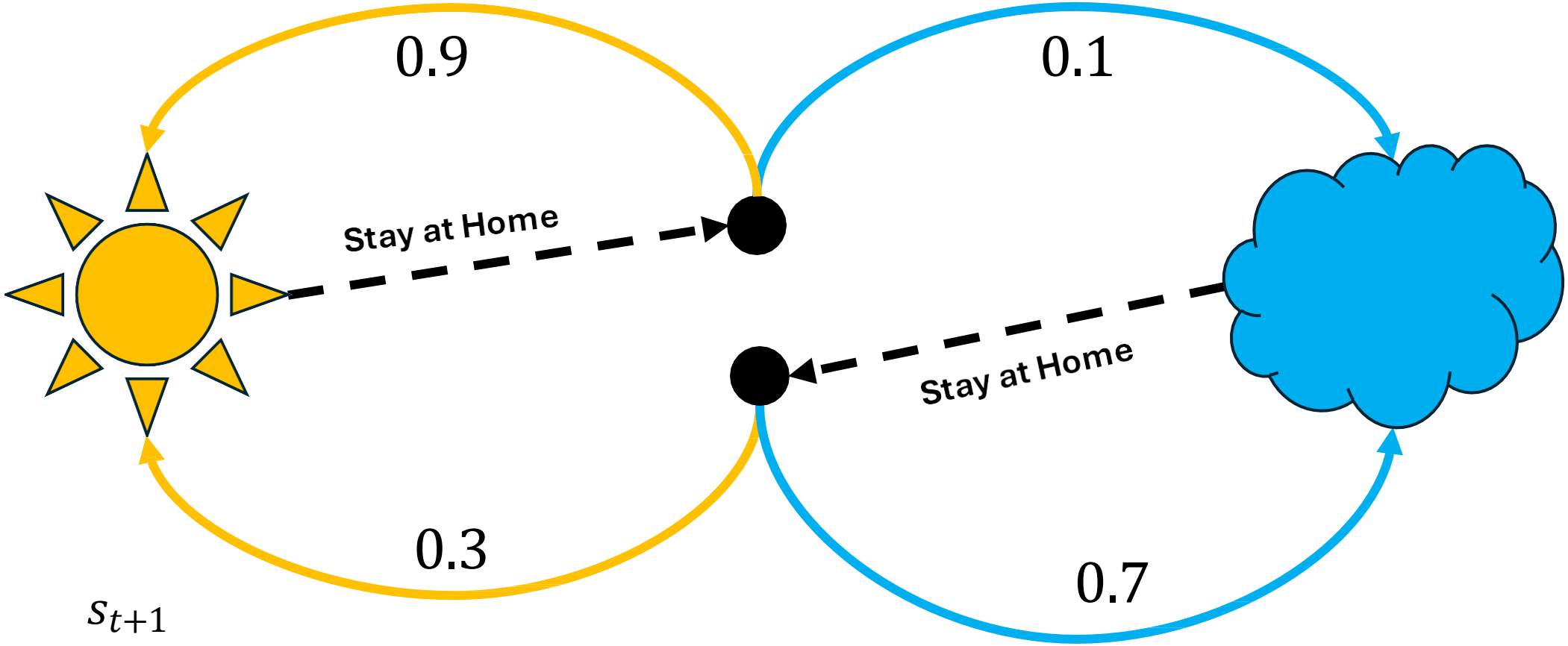


s_{t+1}

s_t			
	0.8	0.2
	0.4	0.6

State Transition Matrix

$$P_{school} = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$



s_{t+1}

s_t		
	0.9	0.1
	0.3	0.7

State Transition Matrix

$$P_{stay} = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$$

Step 1: Compute state-wise average reward under the policy π

For **Sunny**:

$$r_\pi = 0.5(5) + 0.5(-5) = 2.5 + (-2.5) = 0$$

For **Cloudy**:

$$r_\pi = 0.5(3) + 0.5(1) = 1.5 + 0.5 = 2$$

$$\boxed{r_\pi = \begin{bmatrix} 0 \\ 2 \end{bmatrix}}$$

Step 2: Compute the policy transition matrix

Row 1 (Sunny):

- $P_\pi(1,1) = 0.5(0.8) + 0.5(0.9) = 0.4 + 0.45 = 0.85$
- $P_\pi(1,2) = 0.5(0.2) + 0.5(0.1) = 0.1 + 0.05 = 0.15$

Row 2 (Cloudy):

- $P_\pi(2,1) = 0.5(0.4) + 0.5(0.3) = 0.2 + 0.15 = 0.35$
- $P_\pi(2,2) = 0.5(0.6) + 0.5(0.7) = 0.3 + 0.35 = 0.65$

$$\boxed{P_\pi = \begin{bmatrix} 0.85 & 0.15 \\ 0.35 & 0.65 \end{bmatrix}}$$

Step 3: Write the Bellman expectation equations $v_\pi(sunny)$

General Form:

$$v_\pi(s) = r_\pi(s) + \gamma \sum P_\pi(s, s')\, v_\pi(s')$$

$$v_1 = 0 + 0.9(0.85v_1 + 0.15v_2)$$
$$v_1 = 0 + 0.765v_1 + 0.135v_2$$
$$v_1 - 0.765v_1 - 0.135v_2 = 0$$

$$\boxed{0.235v_1 - 0.135v_2 = 0}$$

Step 3: Write the Bellman expectation equations $v_\pi(cloudy)$

General Form:

$$v_\pi(s) = r_\pi(s) + \gamma \sum P_\pi(s, s')\, v_\pi(s')$$

$$v_2 = 2 + 0.9(0.35v_1 + 0.65v_2)$$
$$v_2 = 2 + 0.315v_1 + 0.585v_2$$
$$v_2 - 0.315v_1 - 0.585v_2 = 2$$

$$\boxed{-0.315v_1 + 0.415v_2 = 2}$$

Step 4: Solve for $v_\pi(cloudy)$

$$0.235v_1 - 0.135v_2 = 0$$
$$0.235v_1 = 0 + 0.135v_2$$
$$\frac{0.235v_1}{0.235} = \frac{0 + 0.135v_2}{0.235}$$

$$\boxed{v_1 = \frac{0 + 0.135v_2}{0.235}}$$

$$-0.315v_1 + 0.415v_2 = 2$$
$$-0.315\left(\frac{0 + 0.135v_2}{0.235}\right) + 0.415v_2 = 2$$
$$\left(-0.315 \times \frac{0}{0.235}\right)\left(-0.315 \times \frac{0.135}{0.235}v_2\right) + 0.415v_2 = 2$$
$$(-0.315 \times 0)(-0.315 \times 0.574v_2) + 0.415v_2 = 2$$
$$0 - 0.181v_2 + 0.415v_2 = 2$$
$$0 + 0.234v_2 = 2$$
$$0.234v_2 = 2 + 0$$

$$\boxed{v_\pi(cloudy) = \frac{2 + 0}{0.234} = \frac{2}{0.234} = 8.547}$$

Step 4: Solve for $v_\pi(sunny)$

$$v_1 = \frac{0 + 0.135 v_2}{0.235}$$

$$v_1 = \frac{0 + 0.135(8.547)}{0.235} = \frac{0 + 1.154}{0.235} = \frac{1.154}{0.235}$$

$$v_1(sunny) = 4.911$$

Step 5: Write the Bellman optimality equations

General Form:

$$v_*(s) = \max_a \{R(s,a) + \gamma s' \sum P(s' \mid s,a) v*(s')\}$$

For Sunny (v_1) using School:

$$v_*(sunny) = 5 + 0.9(0.8v_1 + 0.2v_2)$$
$$v_1 = 5 + 0.72v_1 + 0.18v_2$$
$$v_1 - 0.72v_1 - 0.18v_2 = 5$$
$$\boxed{0.28v_1 - 0.18v_2 = 5}$$

For Cloudy (v_2) using School:

$$v_*(cloudy) = 3 + 0.9(0.4v_1 + 0.6v_2)$$
$$v_2 = 3 + 0.36v_1 + 0.54v_2$$
$$v_2 - 0.36v_1 - 0.54v_2 = 3$$
$$\boxed{-0.36v_1 + 0.46v_2 = 3}$$

Step 6: Solve for v_*

From Sunny:

$$0.28v_1 = 5 + 0.18v_2$$

$$v_1 = \frac{5 + 0.18v_2}{0.28}$$

For Cloudy:

$$-0.36v_1 + 0.46v_2 = 3$$

$$-0.36\left(\frac{5 + 0.18v_2}{0.28}\right) + 0.46v_2 = 3$$

Step 6: Solve for v_*(cloudy)

$$-0.36\left(\frac{5+0.18v_2}{0.28}\right)+0.46v_2=3$$

$$\left(-0.36\times\frac{5}{0.28}\right)\left(-0.36\times\frac{0.18v_2}{0.28}\right)+0.46v_2=3$$

$$-6.429-0.231v_2+0.46v_2=3$$

$$-6.429+0.229v_2=3$$

$$0.229v_2=3+6.429$$

$$0.229v_2=9.429$$

$$\boxed{v_*(cloudy)=\frac{9.429}{0.229}=41.175}$$

Step 6: Solve for v_*(sunny)

$$v_1 = \frac{5 + 0.18v_2}{0.28}$$

$$v_1 = \frac{5 + 0.18(41.175)}{0.28}$$

$$v_1 = \frac{5 + 7.412}{0.28} = \frac{12.412}{0.28}$$

$$\boxed{v_*(sunny) = 44.329}$$

Step 7: Solve for q_*

$$q(1, School) = 5 + 0.9(0.8v_1 + 0.2v_2) = 44.33$$
$$q(1, Home) = -5 + 0.9(0.9v_1 + 0.1v_2) = 34.61$$
$$q(2, School) = 3 + 0.9(0.4v_1 + 0.6v_2) = 41.18$$
$$q(2, Home) = 1 + 0.9(0.3v_1 + 0.7v_2) = 38.91$$