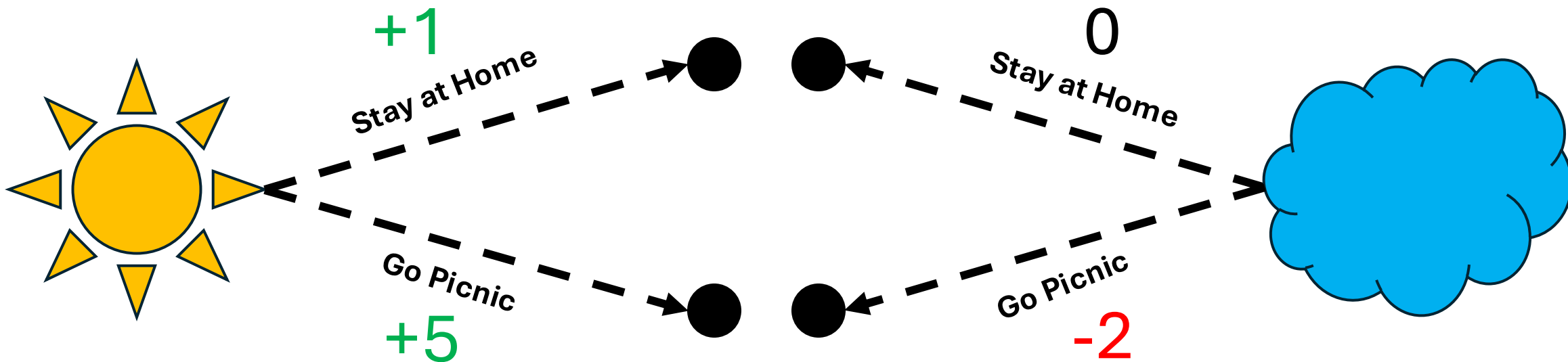



**States**  $s$ : {Sunny, Cloudy}

**Actions**  $a$ : {Go to Picnic, Stay at Home}

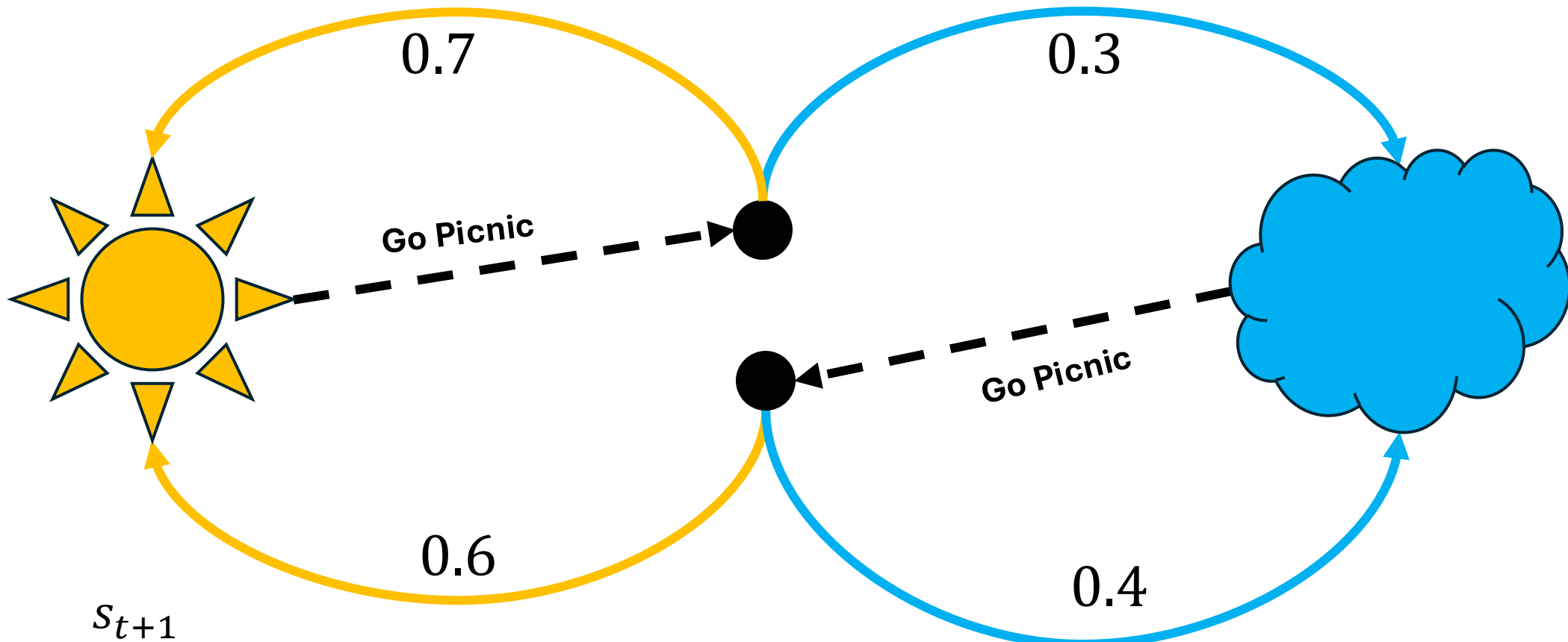
**Discount**  $\gamma = 0.9$





	Go to Picnic	Stay at Home
	+5	+1
	-2	0



	Go to Picnic	Stay at Home
	+5	+1
	-2	0

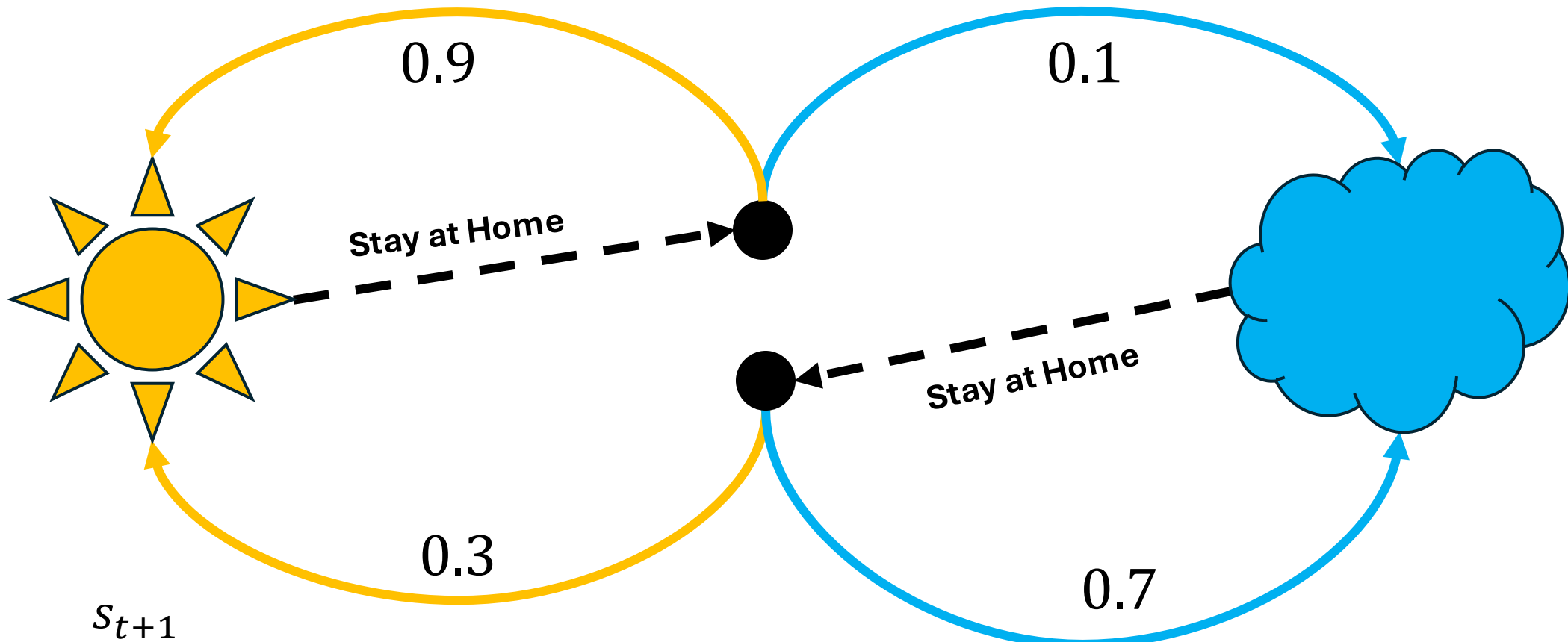
$$R_{picnic} = \begin{bmatrix} 5 \\ -2 \end{bmatrix} \quad R_{stay} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$







		
	0.7	0.3
	0.6	0.4

**State Transition Matrix**

$$P_{picnic} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$$



		
	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 $\pi$

For **Sunny**:

$$r_{\pi} = 0.5 \times (5) + 0.5 \times (1) = 2.5 + 0.5 = 3$$

For **Cloudy**:

$$r_{\pi} = 0.5 \times (-2) + 0.5 \times (0) = -1 + 0 = -1$$

$$r_{\pi} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

That's “what you get right now” on average if you follow the policy

## Step 2: Compute the policy transition matrix

Row 1 (Sunny):

- $P\pi(1,1) = 0.5 \times 0.7 + 0.5 \times 0.9 = 0.35 + 0.45 = 0.80$
- $P\pi(1,2) = 0.5 \times 0.3 + 0.5 \times 0.1 = 0.15 + 0.05 = 0.20$

Row 2 (Cloudy):

- $P\pi(2,1) = 0.5 \times 0.6 + 0.5 \times 0.3 = 0.30 + 0.15 = 0.45$
- $P\pi(2,2) = 0.5 \times 0.4 + 0.5 \times 0.7 = 0.20 + 0.35 = 0.55$

$$P_{\pi} = \begin{bmatrix} 0.80 & 0.20 \\ 0.45 & 0.55 \end{bmatrix}$$

This is how we move between states on average under the policy  $\pi$

# Step 3: Write the Bellman expectation equations $v_{\pi}(\textit{sunny})$

General Form:

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

$$v_1 = 3 + 0.9(0.8 v_1 + 0.2 v_2)$$

$$v_1 = 3 + 0.72 v_1 + 0.18 v_2$$

$$v_1 - 0.72 v_1 - 0.18 v_2 = 3$$

$$0.28 v_1 - 0.18 v_2 = 3$$



## Step 3: Write the Bellman expectation equations $v_\pi(\textit{cloudy})$

General Form:

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

$$v_2 = -1 + 0.9(0.45 v_1 + 0.55 v_2)$$

$$v_2 = -1 + 0.405 v_1 + 0.495 v_2$$

$$v_2 - 0.405 v_1 - 0.495 v_2 = -1$$

$$-0.405 v_1 + 0.505 v_2 = -1$$

Step 4: Write the Bellman expectation equations  $v_{\pi}(\textit{cloudy})$

$$0.28v_1 - 0.18v_2 = 3$$

$$0.28v_1 = 3 + 0.18v_2$$

$$\frac{0.28v_1}{0.28} = \frac{3 + 0.18v_2}{0.28}$$

$$v_1 = \frac{3 + 0.18v_2}{0.28}$$

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

$$[-0.405 v_1 + 0.505 v_2 = -1]$$

$$-0.405\left(\frac{3 + 0.18 v_2}{0.28}\right) + 0.505 v_2 = -1$$

$$\begin{aligned} &(-0.405 \times \frac{3}{.28})(-0.405 \times \frac{.18}{.28} v_2) + 0.505 v_2 \\ &= -1 \end{aligned}$$

$$\begin{aligned} &(-0.405 \times 10.714)(-0.405 \times 0.642857 v_2) + 0.505 v_2 \\ &= -1 \end{aligned}$$

$$-4.339 - 0.2607 v_2 + 0.505 v_2 = -1$$

$$-4.339 + (0.505 - 0.2607) v_2 = -1$$

$$-4.339 + 0.2443 v_2 = -1$$

$$0.2443 v_2 = -1 + 4.339$$

$$v_{\pi}(cloudy) = \frac{-1 + 4.339}{0.2443} = \frac{3.339}{0.2443} = 13.65$$

## Step 4: Solve for $v_{\pi}(\textit{sunny})$

$$v_1 = \frac{3 + 0.18 v_2}{0.28}$$

$$v_1 = \frac{3 + 0.18 (13.65)}{0.28} = \frac{3 + 2.457}{0.28} = \frac{5.457}{0.28}$$

$$v_{\pi}(\textit{sunny}) = 19.489$$

# Step 5: Write the Bellman optimality equations

General Form:

$$v_*(s) = \max_a \{R(s, a) + \gamma \sum_{s'} P(s' | s, a) v_*(s')\}$$

For Sunny ( $v_1$ ) using Picnic:

$$v_*(sunny) = 5 + 0.9(0.7 v_1 + 0.3 v_2)$$

For Cloudy ( $v_2$ ) using Picnic:

$$v_*(cloudy) = -2 + 0.9(0.6 v_1 + 0.4 v_2)$$

# Step 5: Write the Bellman optimality equations

**Sunny:**

$$v_1 = 5 + 0.63 v_1 + 0.27 v_2$$

$$v_1 - 0.63v_1 - 0.27v_2 = 5$$

$$\boxed{0.37v_1 - 0.27v_2 = 5}$$

**Cloudy:**

$$v_2 = -2 + 0.54 v_1 + 0.36 v_2$$

$$\boxed{-0.54 v_1 + 0.64 v_2 = -2}$$

## Step 6: Solve for $v_*$

From the Sunny equation:

$$0.37v_1 = 5 + 0.27 v_2$$

$$v_1 = \frac{5 + 0.27 v_2}{0.37}$$

*Using the equation for cloudy,*

$$\boxed{-0.54 v_1 + 0.64 v_2 = -2}$$

$$-0.54 \left( \frac{5 + 0.27 v_2}{0.37} \right) + 0.64 v_2 = -2$$

## Step 6: Solve for $v_*$ (cloudy)

$$-0.54 \times \frac{5}{0.37} = -4.339$$

$$-0.54 \times \frac{0.27}{0.37} = -0.394 v_2$$

$$-4.339 - 0.394 v_2 + 0.64 v_2 = -2$$

$$-4.339 + (0.64 - 0.394) v_2 = -2$$

$$-4.339 + 0.245 v_2 = -2$$

$$0.245 v_2 = -2 + 4.339 = 5.297$$

$$0.245 v_2 = 5.297$$

$$v_*(\text{cloudy}) = \frac{5.297}{0.245} = 21.538$$



## Step 6: Solve for $v_*$ (sunny)

$$v_1 = \frac{5 + 0.27v_2}{0.37}$$

$$v_1 = \frac{5 + 0.27 \times 21.538}{0.37}$$

$$v_1 = \frac{5 + 5.81}{0.37} = \frac{10.81}{0.37}$$

$$v_*(\text{sunny}) = 29.23$$

$$v_*(\text{cloudy}) = 21.538$$

## Step 7: Solve for $q_*$

- $q(1, \text{Picnic}) = 5 + 0.9(0.7v_1 + 0.3v_2) = 29.23$
- $q(1, \text{Stay}) = 1 + 0.9(0.9v_1 + 0.1v_2) = 26.61$
- $q(2, \text{Picnic}) = -2 + 0.9(0.6v_1 + 0.4v_2) = 21.53$
- $q(2, \text{Stay}) = 0 + 0.9(0.3v_1 + 0.7v_2) = 21.46$

