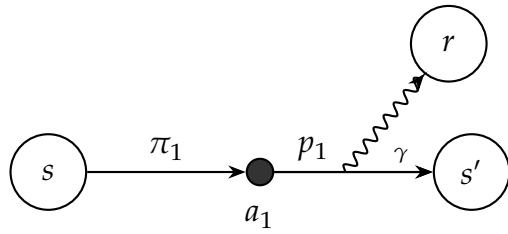


Bellman Equation as Dot Products



$$\begin{aligned}
 g &= \vec{\gamma} \cdot \vec{r} \\
 \vec{o} &= \vec{r} + \gamma \vec{v}' \\
 q &= \vec{p} \cdot \vec{o} &= E [g | s, a] \\
 v &= \vec{\pi} \cdot \vec{q} &= E [g | s]
 \end{aligned}$$

$$\vec{\gamma} = \begin{bmatrix} \gamma^0 \\ \gamma^1 \\ \vdots \end{bmatrix}$$

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} p(s'_1 | a, s) \\ p(s'_2 | a, s) \\ \vdots \end{bmatrix}$$

$$\vec{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} r(s, a, s'_1) \\ r(s, a, s'_2) \\ \vdots \end{bmatrix}$$

$$\vec{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \vec{p}_1 \cdot \vec{o}_1 \\ \vec{p}_2 \cdot \vec{o}_2 \\ \vdots \end{bmatrix}$$

$$\vec{v} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} v(s_0) \\ v(s_1) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vec{\pi}_1 \cdot \vec{q}_1 \\ \vec{\pi}_2 \cdot \vec{q}_2 \\ \vdots \end{bmatrix}$$

$$\vec{\pi} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} p(a_1 | s) \\ p(a_2 | s) \\ \vdots \end{bmatrix}$$

$$\vec{v}' = \begin{bmatrix} v'_0 \\ v'_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} v(s'_0) \\ v(s'_1) \\ \vdots \end{bmatrix}$$

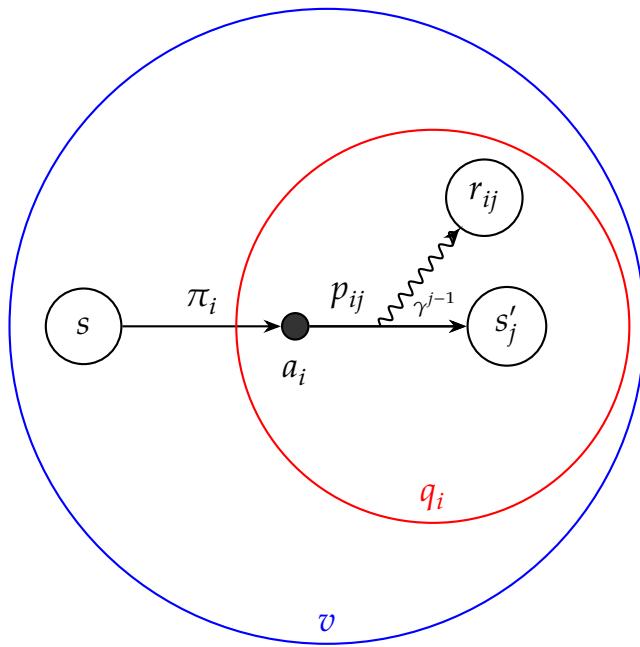
$$\vec{o} = \begin{bmatrix} o_1 \\ o_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} r_1 + \gamma v'_1 \\ r_2 + \gamma v'_2 \\ \vdots \end{bmatrix}$$

Einstein Notation

(implicit summation over repeated indices)

$i = \text{actions}$

$j = \text{next states}$



$$g = \gamma^{j-1} r_j \quad (\text{sum over time steps; over next states})$$

$$o_{ij} = r_{ij} + \gamma v_j' \quad (\text{elementwise — no summation})$$

$$q_i = p_{ij} o_{ij} \quad (\text{sum over next states } j)$$

$$v = \pi_i q_i \quad (\text{sum over actions } i)$$

$$v = \pi_i p_{ij} (r_{ij} + \gamma v_j') \quad (\text{full Bellman equation — sum over } i \text{ and } j)$$

GLOSSARY:

r = reward	(rewarded after action a is taken)
γ = discounting factor	
$g = g_0 = r_1 + \gamma r_2 + \gamma^2 r_3 + \dots$ = return	(reward now plus discounted future reward)
$o = r + \gamma v'$ = outcome	($v = E[g s] = E[o s] \leftarrow$ Recursion: Bellman's insight!)
$s = s_0$ = current state	
$s' = s_1$ = next state	
$v = v(s) = v(s_0)$	
$v' = v(s') = v(s_1)$	
a = action	
p = transition probability, dynamics	
v = state value function	
q = state-action value function	
π = policy	