Bellman Equation for $v_\pi(s)$

Stuff we want to write $v_\pi(s)$ in terms of:

$$\pi(a|s) \stackrel{\text{def}}{=} \Pr(A_t = a|S_t = s)$$
$$p(s', r|s, a) \stackrel{\text{def}}{=} \Pr(S_{t+1} = s', R_{t+1} = r|S_t = s, A_t = a)$$

Partition theorem or law of total expectation:

$$\mathbb{E}[X] = \sum_y \Pr(Y = y)\mathbb{E}\big[X\big|Y = y\big] \tag{1}$$

Definition of $v_\pi(s)$:

$$v_\pi(s) \stackrel{\text{def}}{=} \mathbb{E}_\pi\big[G_t\big|S_t = s\big]$$

Pull one reward out of the return:

$$v_\pi(s) = \mathbb{E}_\pi\big[R_{t+1} + \gamma G_{t+1}\big|S_t = s\big]$$

Apply Equation 1 to condition the expectation on actions:

$$v_\pi(s) = \sum_a \pi(a|s)\mathbb{E}_\pi\big[R_{t+1} + \gamma G_{t+1}\big|S_t = s, A_t = a\big]$$

Split the expectation of a sum into a sum of expectations (note that given an action, the expected immediate reward doesn't depend on the policy):

$$v_\pi(s) = \sum_a \pi(a|s)\big(\mathbb{E}\big[R_{t+1}\big|S_t = s, A_t = a\big] + \gamma\mathbb{E}_\pi\big[G_{t+1}\big|S_t = s, A_t = a\big]\big)$$

Write expected immediate reward in terms of $p(s', r|s, a)$:

$$v_\pi(s) = \sum_a \pi(a|s)\left(\sum_r r\sum_{s'} p(s', r|s, a) + \gamma\mathbb{E}_\pi\big[G_{t+1}\big|S_t = s, A_t = a\big]\right)$$
$$v_\pi(s) = \sum_a \pi(a|s)\left(\sum_{s',r} p(s', r|s, a)r + \gamma\mathbb{E}_\pi\big[G_{t+1}\big|S_t = s, A_t = a\big]\right)$$

Apply Equation 1 to condition the other expectation on the next state:

$$v_\pi(s) = \sum_a \pi(a|s)\left(\sum_{s',r} p(s', r|s, a)r + \gamma\sum_{s',r} p(s', r|s, a)\mathbb{E}_\pi\big[G_{t+1}\big|S_t = s, A_t = a, S_{t+1} = s'\big]\right)$$

By the Markov property, knowing S_{t+1} makes the expectation independent of S_t and A_t:

$$v_\pi(s) = \sum_a \pi(a|s)\left(\sum_{s',r} p(s', r|s, a)r + \gamma\sum_{s',r} p(s', r|s, a)\mathbb{E}_\pi\big[G_{t+1}\big|S_{t+1} = s'\big]\right)$$

Acknowledging that $v_\pi(s') = \mathbb{E}_\pi\big[G_{t+1}\big|S_{t+1} = s'\big]$, and combining the summations:

$$v_\pi(s) = \sum_a \pi(a|s)\left(\sum_{s',r} p(s', r|s, a)r + \gamma\sum_{s',r} p(s', r|s, a)v_\pi(s')\right)$$
$$v_\pi(s) = \sum_a \pi(a|s)\sum_{s',r} p(s', r|s, a)\big(r + \gamma v_\pi(s')\big)$$