where the reward r and next state s' are drawn from a (conditional) probability

distribution $p(s', r|s, a)$, the **Bellman Expectation Equation (for v_π)** expresses the value of any state s in terms of the *expected* immediate reward and the *expected* value of the next state:

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

Calculating the Expectation

In the event that the agent's policy π is **deterministic**, the agent selects action $\pi(s)$ when in state s, and the Bellman Expectation Equation can be rewritten as the sum over two variables (s' and r):

$$v_\pi(s) = \sum_{s' \in \mathcal{S}^+, r \in \mathcal{R}} p(s', r|s, \pi(s))(r + \gamma v_\pi(s'))$$

In this case, we multiply the sum of the reward and discounted value of the next state $(r + \gamma v_\pi(s'))$ by its corresponding probability $p(s', r|s, \pi(s))$ and sum over all possibilities to yield the expected value.

If the agent's policy π is **stochastic**, the agent selects action a with probability $\pi(a|s)$ when in state s, and the Bellman Expectation Equation can be rewritten as the sum over three variables (s', r, and a):

$$v_\pi(s) = \sum_{s' \in \mathcal{S}^+, r \in \mathcal{R}, a \in \mathcal{A}(s)} \pi(a|s) p(s', r|s, a)(r + \gamma v_\pi(s'))$$

In this case, we multiply the sum of the reward and discounted value of the next state $(r + \gamma v_\pi(s'))$ by its corresponding probability $\pi(a|s)p(s', r|s, a)$ and sum over all possibilities to yield the expected value.

There are 3 more Bellman Equations!

In this video, you learned about one Bellman equation, but there are 3 more, for a total of 4 Bellman equations.