In reinforcement learning, sequential decision making problems need to be solved by an agent. The agent is provided with some input state and then needs to choose an action. Input state and chosen action are used to create a subsequent state and a reward signal for the agent. By maximizing the reward signals the agent develops a policy for acting in the environment and thus gets better in solving the given problem. Finding the optimal policy is the goal of reinforcement learning. Hereby the environment is fully characterized by the state.
Time is discrete in reinforcement learning. Therefore the consecutive states form a chain. Nevertheless transitions from one state to its subsequent state are not dependent on history, they only depend on the current state. This is called the Markov assumption.
For finding the optimal policy the state value function $V_\pi(s)$ can be used. It is defined as the expected reward of a policy π that can be reached by starting at the state s

$$V_\pi = \mathbb{E}\left[R\right]. \tag{1}$$

The expected reward R for a given state s is defined as the sum of the current reward and the discounted future rewards

$$R = \sum_{t=0}^{\infty} \gamma^t \, R_t \text{ for } s = s_0. \tag{2}$$

$\gamma \, \epsilon \, [0, 1]$ is the discount factor.
The optimal policy π^* maximizes the state-value function.