- If $\gamma = 1$, the return is not discounted.
- For larger values of γ, the agent cares more about the distant future. Smaller values of γ result in more extreme discounting, where - in the most extreme case - agent only cares about the most immediate reward.

MDPs and One-Step Dynamics

- The **state space** $\mathcal{S}$ is the set of all (*nonterminal*) states.
- In episodic tasks, we use $\mathcal{S}^+$ to refer to the set of all states, including terminal states.

- The **action space** $\mathcal{A}$ is the set of possible actions. (Alternatively, $\mathcal{A}(s)$ refers to the set of possible actions available in state $s \in \mathcal{S}$.)
- (Please see **Part 2** to review how to specify the reward signal in the recycling robot example.)
- The **one-step dynamics** of the environment determine how the environment decides the state and reward at every time step. The dynamics can be defined by specifying $p(s', r|s, a) \doteq \mathbb{P}(S_{t+1} = s', R_{t+1} = r|S_t = s, A_t = a)$ for each possible $s', r, s, \text{and } a$.
- A **(finite) Markov Decision Process (MDP)** is defined by:
  - a (finite) set of states $\mathcal{S}$ (or $\mathcal{S}^+$, in the case of an episodic task)
  - a (finite) set of actions $\mathcal{A}$
  - a set of rewards $\mathcal{R}$
  - the one-step dynamics of the environment
  - the discount rate $\gamma \in [0, 1]$