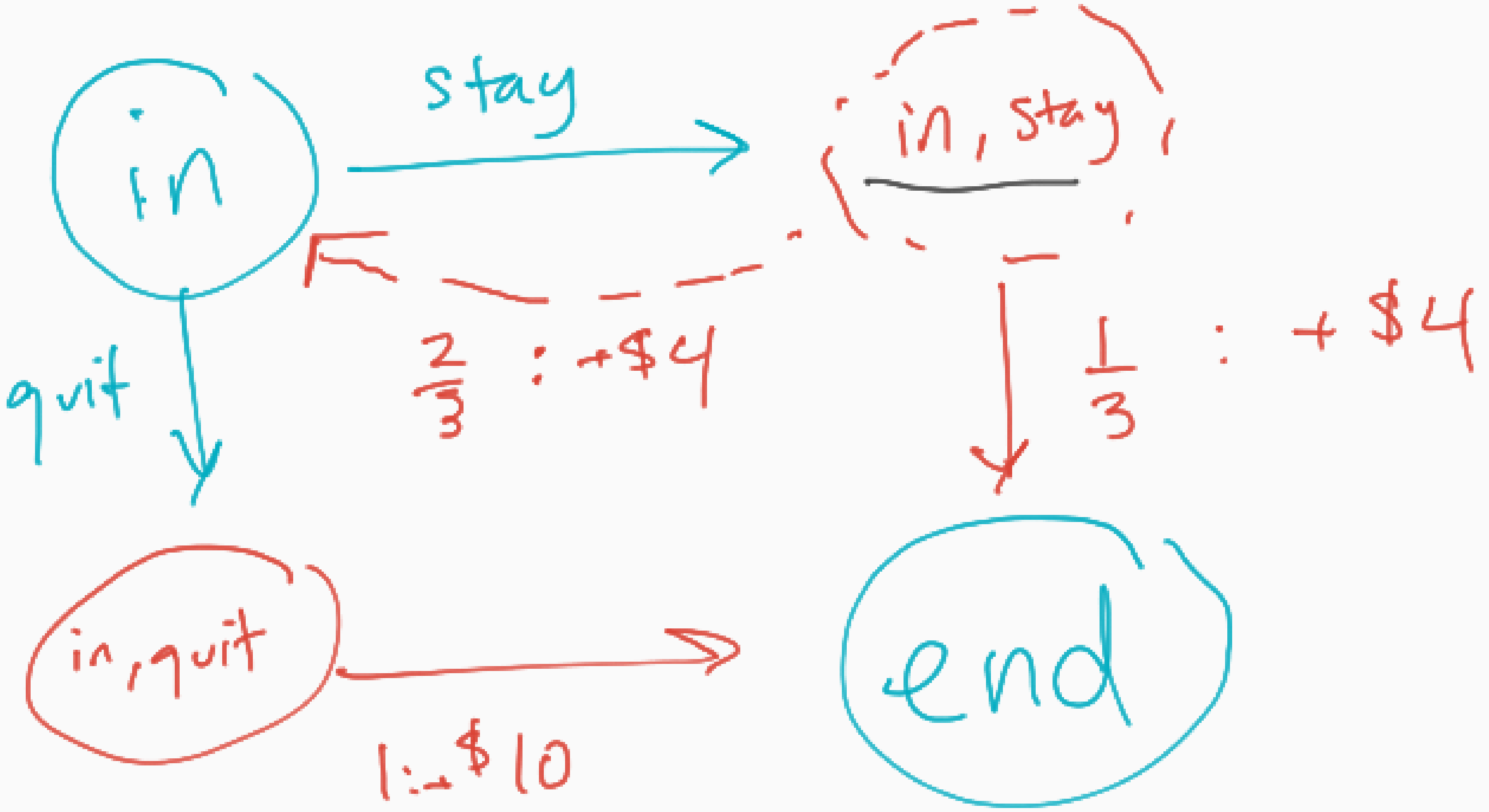


states s

chance nodes

actions a

transition $\underline{T(s' | \underline{s}, a)} \in \mathbb{R}$

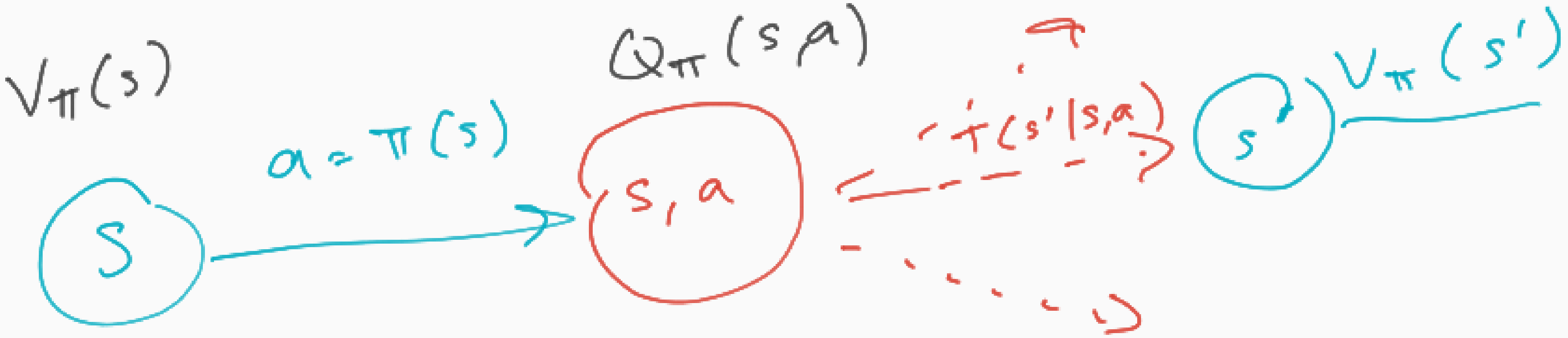
reward $R(\underline{s, a, s'})$

discount factor $0 \leq \gamma \leq 1$

policy $\pi(s) \to a$

utility $\sum_{t=0}^{T} \gamma^t r_t$

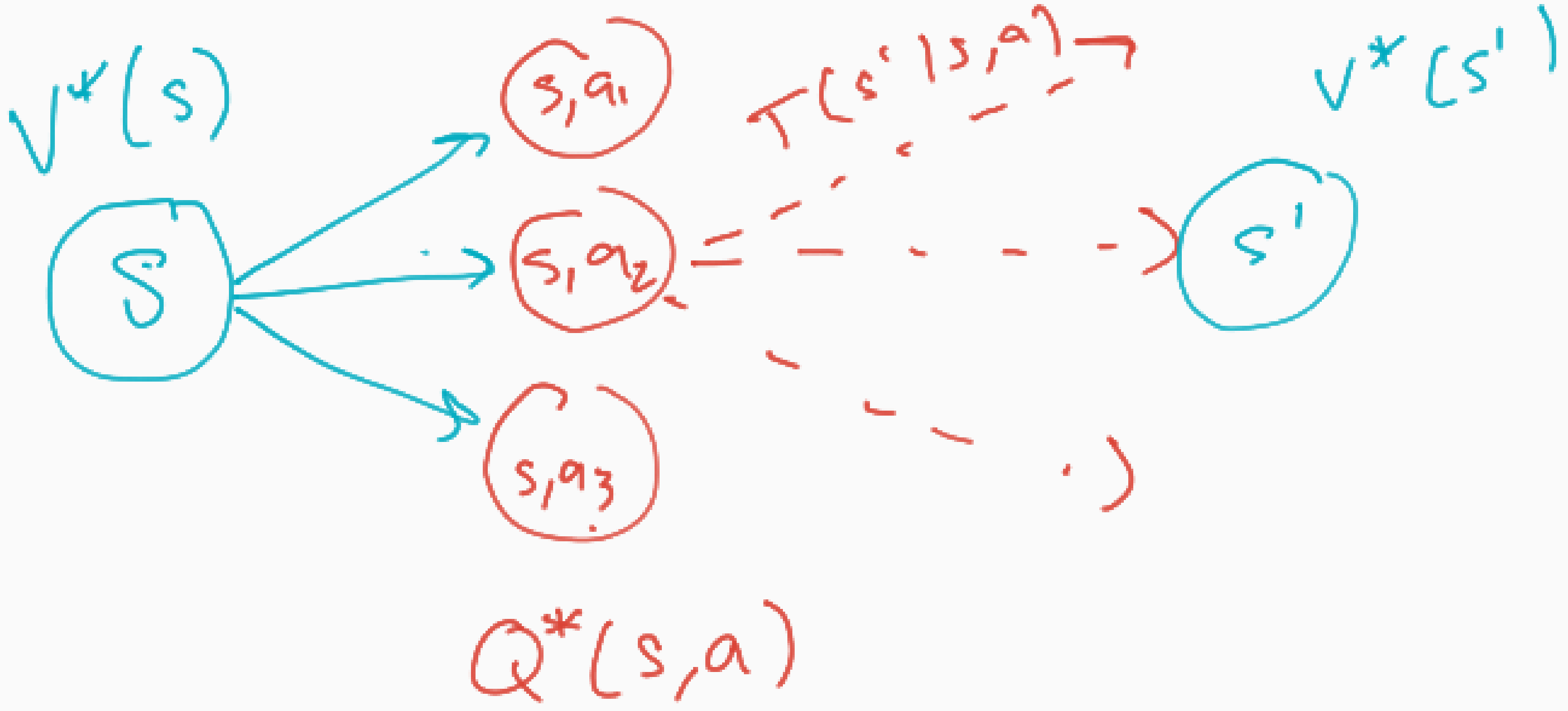
Value $E[U] = V_\pi(s)$



$$\underline{V_\pi(s)} = \begin{cases} 0 & \text{IsEnd}(s) = \text{True} \\ Q_\pi(s, \pi(s)) \end{cases}$$

$$Q_\pi(s,\underline{a}) = \sum_{s'} T(s' | s,a) \left[R(s,a,s') + \underbrace{\gamma V_\pi(s')}_{\gamma Q_\pi(s', \pi(s'))} \right]$$

Bellman Equation



$$\left.\begin{array}{l} V^*(s) = \begin{cases} 0 & \text{isEnd}(s) = \text{True} \\ \max\limits_{a \in \text{Actions}(s)} Q^*(s,a) \end{cases} \\ \\ Q^*(s,a) = \sum\limits_{s'} T(s'|s,a)\left[R(s,a,s') + \gamma V^*(s')\right] \end{array}\right\} \text{Bellman Optimality Equation.}$$

$$\pi^*(s) = \operatorname*{argmax}_{a \in \text{Actions}(s)} Q^*(s,a)$$