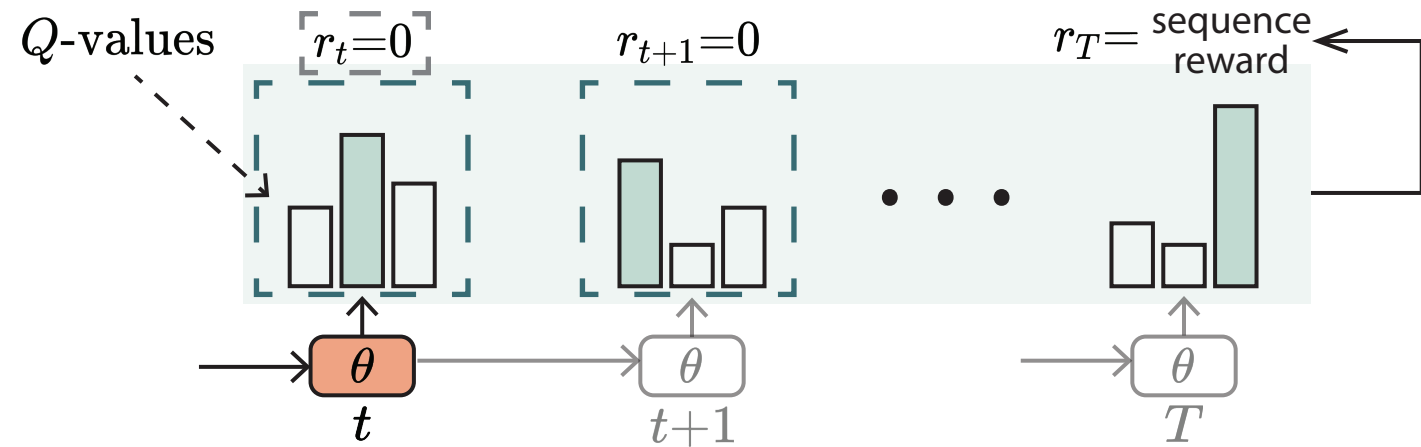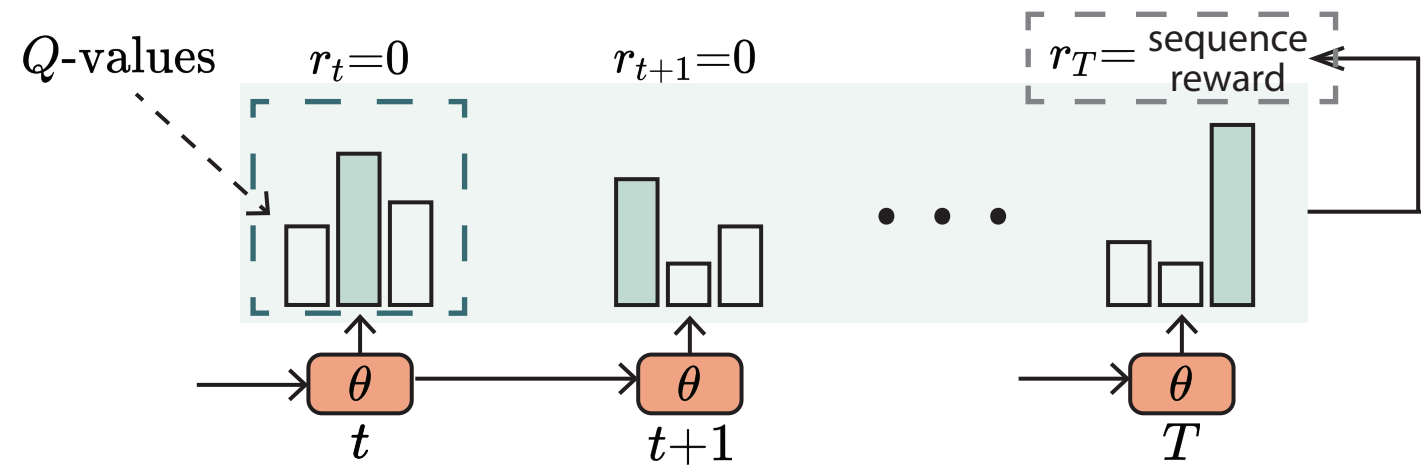$$\mathcal{L}_{\text{SQL, PCL}}(\boldsymbol{\theta}) = \mathbb{E}_{\pi'}\left[\frac{1}{2}\left(-V_{\bar{\theta}}(\boldsymbol{s}_t) + \gamma V_{\bar{\theta}}(\boldsymbol{s}_{t+1}) + r_t - \log \pi_{\theta}(a_t \mid \boldsymbol{s}_t)\right)^2\right]$$

$$\mathcal{L}_{\text{SQL, PCL-ms}}(\boldsymbol{\theta}) = \mathbb{E}_{\pi'}\left[\frac{1}{2}\left(-V_{\bar{\theta}}(\boldsymbol{s}_t) + \gamma^{T-t}r_T + \sum_{l=0}^{T-t}\gamma^l \log \pi_{\theta}(a_{t+l} \mid s_{t+l})\right)^2\right]$$

Single-Step PCL Training

Multi-Step PCL Training