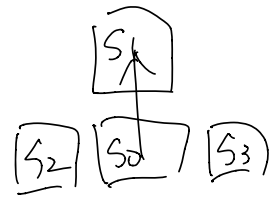
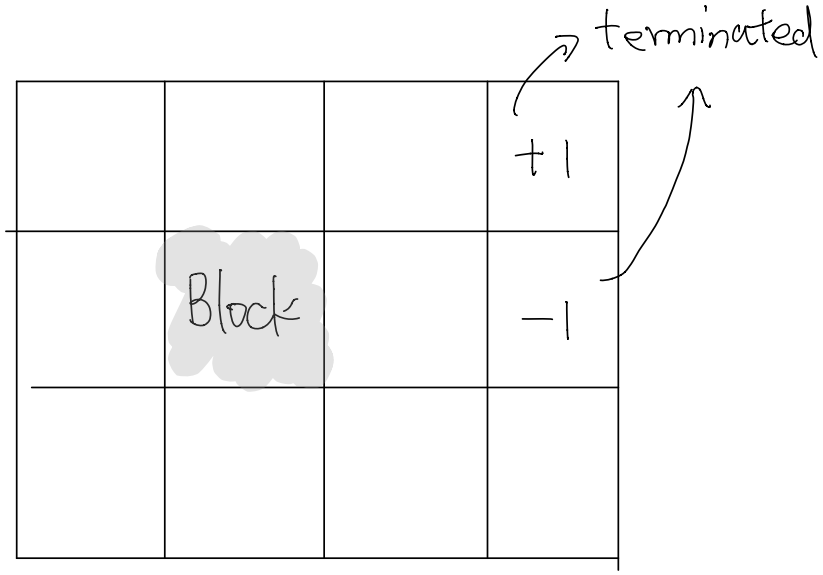
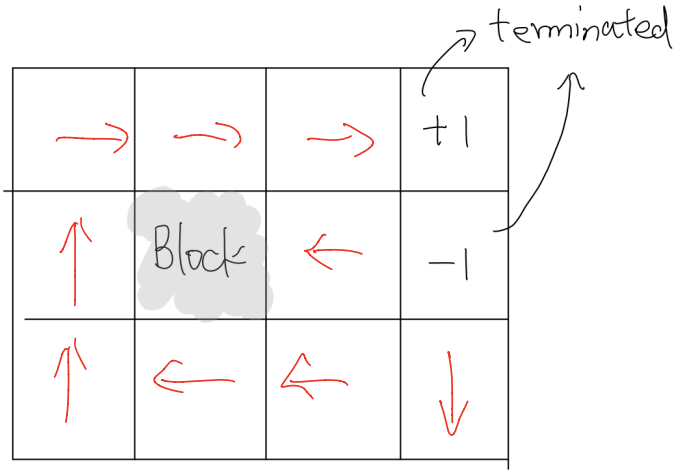
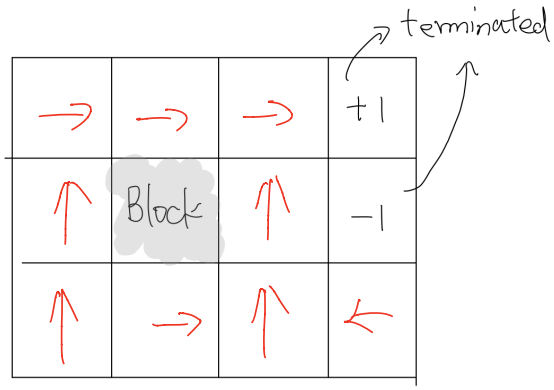
对于每个点，方向不是确定的.



up个 → 0.8到 S_1
→ 0.1到 S_2
→ 0.1到 S_3.



When gamma = 1. Reward = -0.02.



When gamma = 1, Reward = -0.1