

Rewards and discounting

- Reward Function:

$$R(T) = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4}$$

Gamma: discount rate

Cumulative

Reward

Trajectory, Sequence
of States and Actions

$$R(T) = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

- Discounting Rewards (the rewards that come sooner (at the beginning of the game) are more likely to happen since they are more predictable than the long-term future reward)

1. We define a discount rate called gamma. It must be between 0 and 1. Most of the time between 0.99 and 0.95

- The larger the gamma, the smaller the discount. This means that our agent cares more about the long-term reward.

- The smaller the gamma, the bigger the discount. This means that our agent cares more about the short-term reward.

2. Then, each reward will be discounted to the extent of the time step. As time step increases the future reward is less and less likely to happen.