# Part 1: Optimal Policy

The return in this system is computed as: $$ G_t = R_{t+1} + \gamma \cdot R_{t+2} $$ where:

- If the agent chooses the **left action**, $R_{t+1} = 1$ and $R_{t+2} = 0$.
- If the agent chooses the **right action**, $R_{t+1} = 0$ and $R_{t+2} = 2$.

We analyze which policy is optimal under different values of the discount factor $\gamma$:

---

## 1. $\gamma = 0$

When $\gamma = 0$, the return simplifies to just the immediate reward:
$G_t = R_{t+1}$

- **Left action:** $G_t = 1$
- **Right action:** $G_t = 0$

**Optimal Policy:**
In this case, the optimal policy is to always go **left** ($\pi_{left}$), since it gives the higher immediate reward.

---

## 2. $\gamma = 0.9$

Here, the return takes into account both immediate and future rewards: $G_t = R_{t+1} + 0.9 \cdot R_{t+2}$

- **Left action:** $G_t = 1 + 0.9 \cdot 0 = 1$
- **Right action:** $G_t = 0 + 0.9 \cdot 2 = 1.8$

**Optimal Policy:**
The **right action** ($\pi_{right}$) becomes optimal, because the discounted future reward outweighs the immediate reward from going left.

---

## 3. $\gamma = 0.5$

With $\gamma = 0.5$, both policies result in the same total reward:

- **Left action:** $G_t = 1 + 0.5 \cdot 0 = 1$
- **Right action:** $G_t = 0 + 0.5 \cdot 2 = 1$

**Optimal Policy:**
**Either policy** ($\pi_{left}$ or $\pi_{right}$) is optimal, since both yield the same total return.

---

| Discount Factor ($\gamma$) | Left Policy ($G_t$) | Right Policy ($G_t$) | Optimal Policy |
|---|---|---|---|
| 0 | 1 | 0 | Left |
| 0.9 | 1 | 1.8 | Right |

| Discount Factor ($\gamma$) | Left Policy ($G_t$) | Right Policy ($G_t$) | Optimal Policy |
| --- | --- | --- | --- |
| 0.5 | 1 | 1 | Either (Left/Right) |