**Question on Markov Decision Process (MDP)**

**Scenario:** A student is preparing for a **Mathematics Exam**. Each day, the student can choose one of the following study methods:

1. **Self-Study (S)**
2. **Attending Coaching Class (C)**
3. **Watching Online Tutorials (O)**

The student’s performance (marks improvement) depends on the chosen study method, and the process follows a **Markov Decision Process (MDP)** with the following **transition probabilities and rewards**:

**Transition Probabilities:**

* **Self-Study (S):**
  + 60% chance of moving to a **higher level of understanding**.
  + 40% chance of staying at the same level.
* **Coaching Class (C):**
  + 80% chance of moving to a **higher level of understanding**.
  + 20% chance of staying at the same level.
* **Online Tutorials (O):**
  + 50% chance of moving to a **higher level of understanding**.
  + 50% chance of staying at the same level.

**Reward System:**

* **If the student moves to a higher level of understanding:**
  + Self-Study: **+4 marks**
  + Coaching Class: **+6 marks**
  + Online Tutorials: **+3 marks**
* **If the student remains at the same level:**
  + Self-Study: **+1 mark**
  + Coaching Class: **+2 marks**
  + Online Tutorials: **+1 mark**

**Question:**

1. **Model this problem as an MDP** by defining:
   * **States (S)**, **Actions (A)**, **Transition Probabilities (P)**, and **Rewards (R)**.
2. **If the student starts with 50 marks and follows this study plan for 2 days:**
   * **Day 1:** Chooses **Self-Study (S)**
   * **Day 2:** Chooses **Coaching Class (C)**

**What is the expected total marks after 2 days?**

**Solution Outline:**

1. **Define the MDP elements:**
   * **States (S):** {Low Understanding, High Understanding}
   * **Actions (A):** {Self-Study (S), Coaching Class (C), Online Tutorials (O)}
   * **Transition Probabilities (P):**
     + **Self-Study (S):** 60% chance to improve, 40% chance to remain.
     + **Coaching Class (C):** 80% chance to improve, 20% chance to remain.
     + **Online Tutorials (O):** 50% chance to improve, 50% chance to remain.
   * **Rewards (R):**
     + **Self-Study (S):** +4 if improved, +1 if same.
     + **Coaching Class (C):** +6 if improved, +2 if same.
     + **Online Tutorials (O):** +3 if improved, +1 if same.
2. **Expected Marks Calculation (Starting at 50 Marks):**
   * **Day 1 (Self-Study - S)**
     + 60% chance of **improving (50 + 4 = 54)**
     + 40% chance of **staying the same (50 + 1 = 51)**
     + **Expected Marks after Day 1:** (0.6×54)+(0.4×51)=32.4+20.4=52.8(0.6 \times 54) + (0.4 \times 51) = 32.4 + 20.4 = 52.8
   * **Day 2 (Coaching Class - C)**
     + 80% chance of **improving (52.8 + 6 = 58.8)**
     + 20% chance of **staying the same (52.8 + 2 = 54.8)**
     + **Expected Marks after Day 2:** (0.8×58.8)+(0.2×54.8)=47.04+10.96=58(0.8 \times 58.8) + (0.2 \times 54.8) = 47.04 + 10.96 = 58

**Final Answer:**

**The expected total marks after 2 days = 58 marks.**

**Introduction to Markov Decision Process (MDP)**

A **Markov Decision Process (MDP)** is a mathematical framework used for decision-making in situations where the outcomes are partly random and partly controlled by a decision-maker. It is widely used in artificial intelligence, robotics, and economics to model problems involving sequential decision-making.

**Key Components of an MDP**

1. **States (S):** The different situations or conditions in which an agent can be.
2. **Actions (A):** The possible choices or moves an agent can take in each state.
3. **Transition Probability (P):** The probability of moving from one state to another after taking an action.
4. **Reward (R):** The feedback or score the agent gets for performing an action in a state.
5. **Policy (π):** A strategy that defines which action to take in each state.
6. **Discount Factor (γ):** A value between 0 and 1 that represents the importance of future rewards.

**1. Policy (π) – The Study Strategy**

A **policy (π)** is a strategy that tells the student what action to take in each state to maximize their total expected marks over time. In simple words, it is a **decision-making rule** for choosing study methods.

**Example in the Student Scenario:**

* The student starts with **50 marks** and can choose between three study strategies:
  1. **Self-Study (S)**
  2. **Coaching Class (C)**
  3. **Online Tutorials (O)**
* A possible **policy (π)** could be:
  1. If the student's understanding is **low**, choose **Coaching Class (C)** (because it has the highest probability of improving marks).
  2. If the student's understanding is **medium**, choose **Self-Study (S)** (because it provides a decent improvement and is cost-effective).
  3. If the student's understanding is **high**, choose **Online Tutorials (O)** (to revise and maintain marks).

Thus, the **policy (π)** guides the student in selecting the best study method based on their current marks and understanding.

**2. Discount Factor (γ) – The Importance of Future Marks**

The **discount factor (γ)** determines how much future rewards (marks) matter compared to immediate rewards. It is a value between **0 and 1**.

* **If γ is close to 1:**
  + The student **values future marks highly** and focuses on long-term learning rather than quick improvements.
  + They might prefer a strategy like **Self-Study + Coaching Class** over time because it provides **steady improvement**.
* **If γ is close to 0:**
  + The student **only cares about immediate marks** and does not plan long-term.
  + They might always choose **Coaching Class** since it gives the highest short-term gain, even if it’s expensive.

**Example in the Student Scenario:**

* Suppose **γ = 0.9** (high discount factor):
  + The student **plans for long-term learning** and may first choose **Coaching Class**, then shift to **Self-Study** as they gain understanding.
* Suppose **γ = 0.3** (low discount factor):
  + The student **only focuses on short-term results**, preferring Coaching Class repeatedly without considering if it is sustainable.

**Summary:**

* **Policy (π):** A strategy that tells the student which study method to choose in each state.
* **Discount Factor (γ):** Determines how much future marks matter compared to immediate marks.

Thus, a **well-defined policy and a suitable discount factor** help the student **maximize total marks over time** rather than just improving marks in a single day!