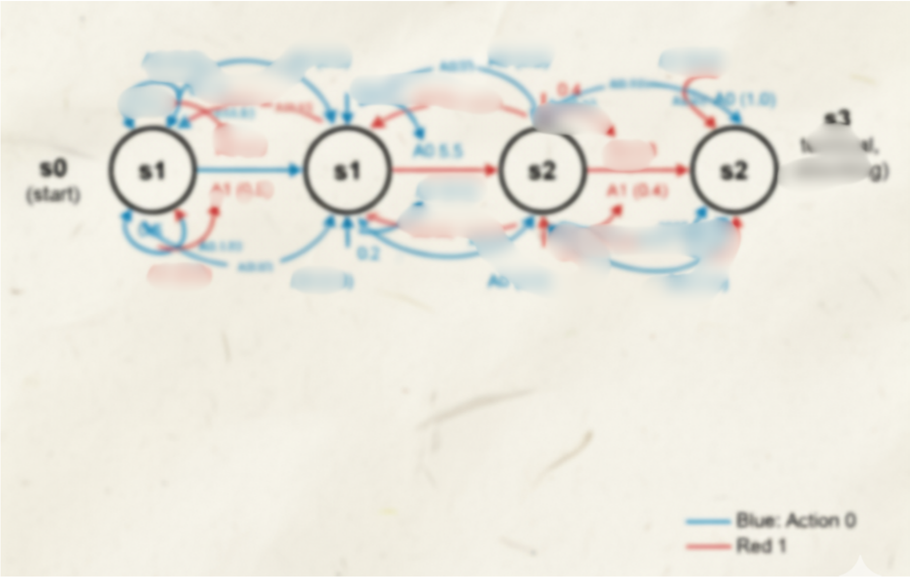
Lab 2 — Getting Familiar with MDPs (Sample Submission)

1) Time Estimate (5 pts)

- **Before starting (estimate):** ~2 hours 15 minutes

5) MDP2 Multi-Graph Embedding (25 pts)

A clear multi-graph shows **the same nodes** with **two edge sets** (one per action). Below is a tidy ASCII version students can copy; colors are replaced by labels A0: and A1:



(Note: This graph is intentionally blurred and provided only as a sample to show the expected structure and formatting of your submission.)

Nodes: s0 (start), s1, s2, s3 (terminal, absorbing)

A0 edges (Action 0):

s0 --X--> s0   s0 --X--> s1

s1 --X--> s0   s1 --X--> s1   s1 --X--> s2

s2 --X--> s1   s2 --X--> s2   s2 --X--> s3

s3 --1.0--> s3 (absorbing)

A1 edges (Action 1):

s0 --X--> s0   s0 --X--> s1

s1 --X--> s0   s1 --X--> s1   s1 --X--> s2

s2 --X--> s1   s2 --X--> s2   s2 --X--> s3

s3 --1.0--> s3 (absorbing)

(Here "X" means students must fill in probabilities from the MDP2 file.)

6) A Reasonable Policy for MDP2 (15 pts)

Goal: maximize expected reward by reaching s3 (the only rewarding state).

Greedy one-step-look policy (intuitive and effective):

- **π(s0) = Action ? (?? to s1 vs ?? under other action)**
- **π(s1) = Action ? (?? to s2 vs ?? under other action)**
- **π(s2) = Action ? (?? to s3 vs ?? under other action)**
- **π(s3) = (either; terminal/absorbing)**

Table form:

State	**Chosen Action**
s0	**?**
s1	**?**
s2	**?**
s3	**— (terminal)**

8) Design Your Own MDP with a "Gadget" (50 pts)

Domain: Robot Vacuum in a Hallway of Rooms

Intuition: Start with a **1-room gadget** (clean or dirty), then "tile" it to make a multi-room hallway.

8.1 One-Room Gadget

- **States:**

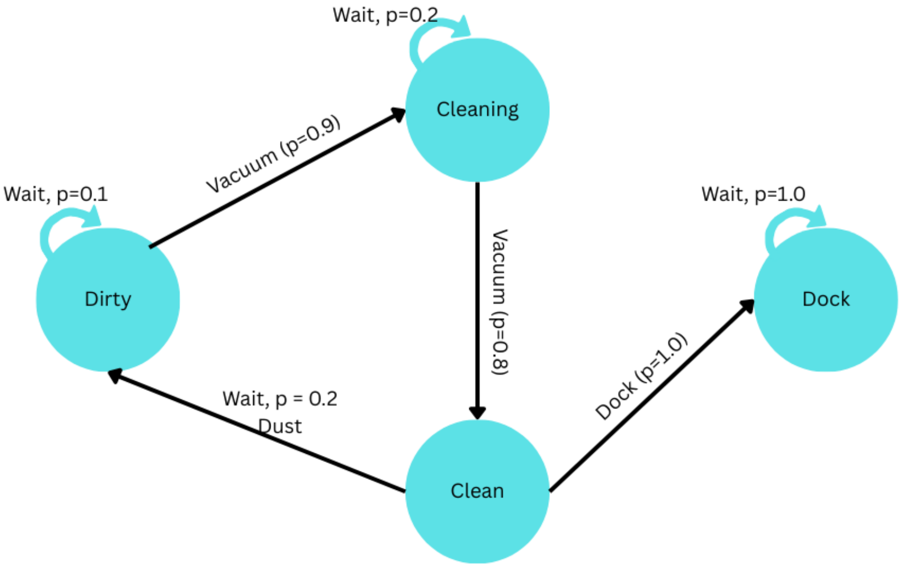
  Dirty, Cleaning, Clean, Dock (terminal)

- **Actions:**

  Vacuum (from Dirty → Cleaning → Clean), Wait, Dock

- **Transitions (sketch):**
  - From Dirty: Vacuum → Cleaning (p=0.9) else stays Dirty (p=0.1)
  - From Cleaning: Vacuum → Clean (p=0.8) else stays Cleaning (p=0.2)
  - From Clean: Dock → Dock (p=1.0); Wait risks dust reappearing: Clean→Dirty (p=0.2)
- **Rewards:**

  +1 upon entering Clean, −0.05 per action as battery cost, +0.5 bonus for entering Dock.



+1 upon entering Clean
+0.5 bonus for entering Dock
-0.05 per action (battery cost)

8.2 Gadget → Hallway (N rooms)

- **Idea:** Chain N copies: (Room_i, status ∈ {Dirty, Cleaning, Clean}) plus shared Dock.
- **New Action:** MoveRight (from Room_i to Room_{i+1}) and MoveLeft to backtrack; small slip p=0.05.
- **Policy idea:** Clean current room → move right; if battery low, head toward Dock.
- **Why it's MDP-worthy:** Local cleaning dynamics **repeat as a gadget**, and the larger hallway trades off **progress vs. battery** with stochastic slips and re-dirty risk.

9) Final Time Report 5 pts)

- **Estimate:** 2h15m
- **Actual:** 2h40m

10) What to Submit (as shown by this sample)

- **One readable file (PDF or DOCX)** containing:
    1. Your MDP2 multi-graph (Sec. 5)
    2. Your policy for MDP2 (Sec. 6)
    3. Your designed MDP + gadget explanation (Sec. 8)
    4. Time estimate vs. actual (Secs. 1 & 9)

(This sample intentionally shows: neat sectioning, a legible multi-graph, a concise policy table, and a clear "gadget → larger system" modeling step—exactly what scorers look for.

- Students may hand-draw the graph (phone scan is fine) or diagram it digitally.