

6.4110
Representation, Inference and Reasoning in AI

Quiz 3

April 27, 2026

Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

You are permitted to use a single sheet of paper with notes on (both sides). You may not use a calculator. **Box all answers** for free response questions.

Name: _____

MIT email: _____

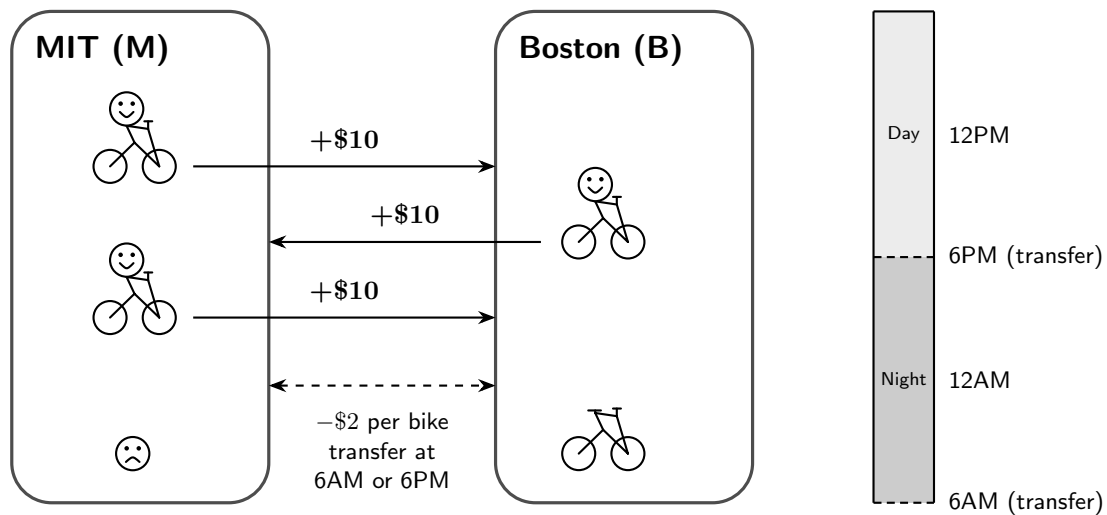
Question	Points	Score
1	23	
2	20	
3	22	
4	26	
5	9	
Total:	100	

1 Blue Bikes

1. You are in charge of a small-city deployment of a bike-share system.

- There are four (4) bikes.
- There are two locations: MIT (M) and Boston (B). Each location has a dock that can contain at most four (4) bikes.
- At 6AM and at 6PM, you can (instantaneously!) move as many bikes as you want between locations.
- During the day (between 6AM and 6PM), and again during the night (between 6PM and 6AM), some number between 0 and 4 (inclusive) people want to ride between the two locations: the number is drawn from a distribution P_{ijk} where i and j are locations (M or B) and $k = D$ for daytime and $k = N$ for nighttime. So, for example, $P_{MBD}(3)$ is the probability that three people want to ride from MIT to Boston during the day. Each bike can be ridden at most once per day and once per night. No person can bike from MIT to MIT or from Boston to Boston directly.
- If someone successfully completes a ride, you earn +10. This can only happen if there is a bike available at the passenger's starting location when they want to ride.
- It costs -2 for the bike-sharing company to move a bike between locations at 6AM or 6PM.

The diagram below illustrates a sample arrangement (not necessarily describing any of the following problems) in which there are two bikes at MIT and two bikes in Boston. Over the course of a day, three people at MIT wish to travel to Boston, while one person in Boston wants to go to MIT. However, there are only enough bicycles for two of the three people at MIT to ride to Boston since the bike from the person in Boston can not be used again until the next night. During the night, assuming no transfers have taken place, there will be one bike at MIT and three bikes in Boston.



Your goal is to move bikes around to maximize your profit! Since you're a 6.4110 student, you decide to formulate this as an MDP.

6.4110

Quiz 3, April 27, 2026

- (a) (5 points) What is the smallest state space we could use to represent this problem? How big is it?

- (b) (5 points) What is the action space (the set of actions that could possibly be taken in any state)?

- (c) (5 points) Assume that today, all the bikes start at MIT at 6AM. If you don't take any action, what is the probability that the system transitions to a state in which there are two bikes at MIT and two at Boston at 6PM? Answer in terms of the appropriate P_{ijk} values.

(d) For each question below, indicate whether it is best modeled as an MDP, a POMDP, a deterministic reward-maximization problem, or another type of problem. If you select “Other,” explain in the box why it cannot be modeled as an MDP, POMDP, or Deterministic problem; if you select any of the other options, you may leave the box blank.

i. (2 points) The same 4 students bike from MIT to Boston during the day, and then bike back from Boston to MIT at midnight.

MDP POMDP Deterministic Other

ii. (2 points) Your bike-dock system is kind of flakey and it sometimes mis-reports the locations of the bikes.

MDP POMDP Deterministic Other

iii. (2 points) Sometimes when a rider tries to take a bike out of the dock, it fails. Assume each rider only tries once before giving up, and failure probabilities are independent.

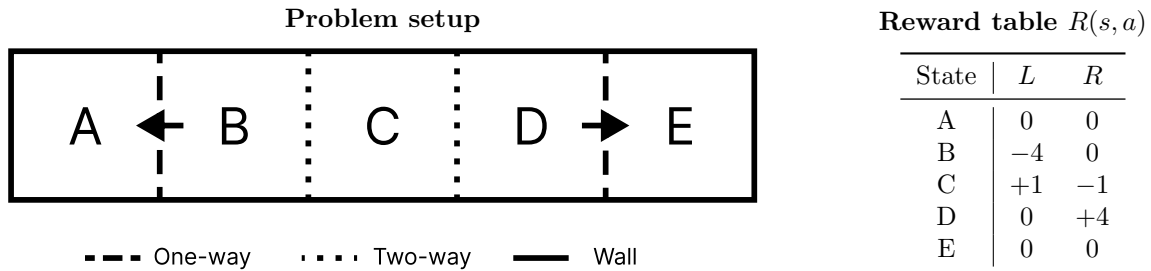
MDP POMDP Deterministic Other

iv. (2 points) Riders tell their friends about where they have had success finding a bike, and their friends change their commuting habits accordingly.

MDP POMDP Deterministic Other

2 MDP Decision-Making

2. Consider the following problem with 5 states $\{A, B, C, D, E\}$ and two actions $\{L, R\}$. States A and E are terminal states; after either one is reached, it cannot be escaped, and all actions taken from those states have 0 reward. In the other states, each action has a $5/6$ chance of moving in the targeted direction; otherwise, you move in the opposite direction. You are given rewards for taking specific actions in specific states (regardless of the actual direction you move). A visualization of the setup is shown on the left below, and a table of the rewards is shown on the right. Use a discount factor of $\gamma = 0.9$.



- (a) (10 points) Assuming a discount factor of 0.9, what are the horizon 0, 1 and 2 value functions? **Fill in the missing rows.** We left some space for calculations at the bottom of this page.

	A	B	C	D	E
$V_0(s)$					
$V_1(s)$					
$V_2(s)$					
$V_3(s)$	0	$\frac{60}{40}$	$\frac{89}{40}$	$\frac{172}{40}$	0

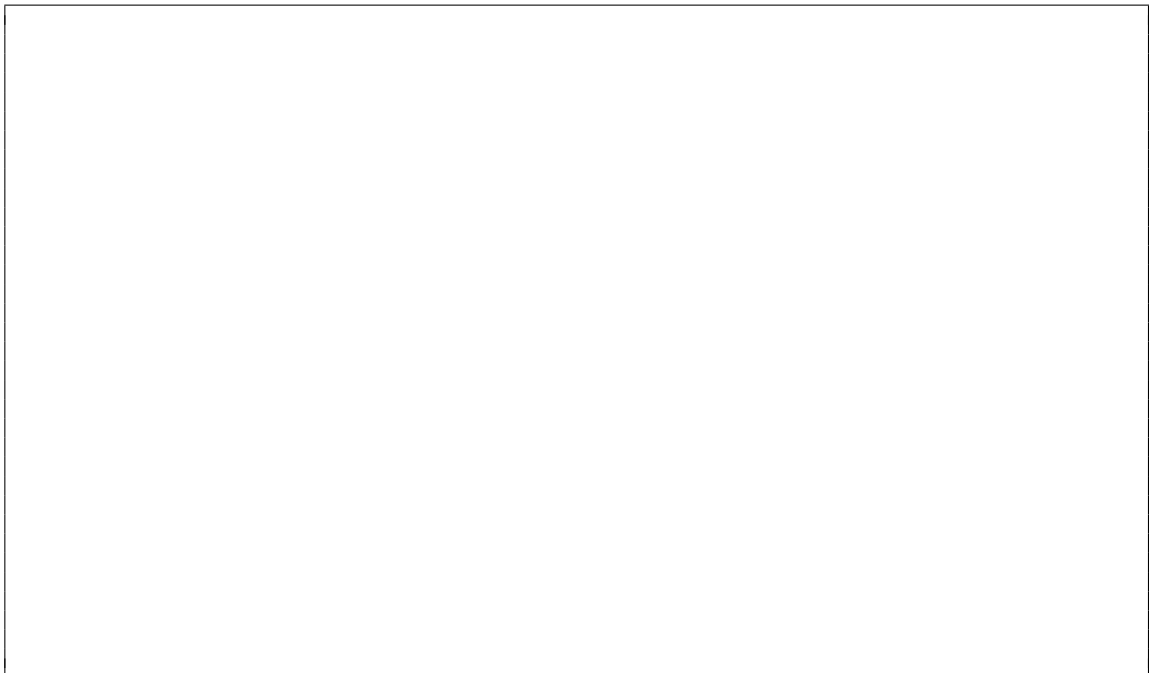
6.4110

Quiz 3, April 27, 2026

- (b) (5 points) Draw the *policy tree* for the optimal policy when starting in state 3 using the horizon 3 values.



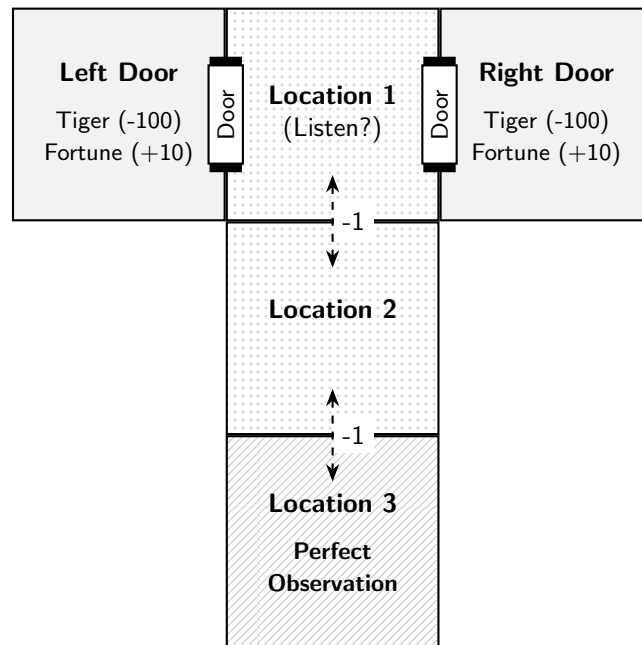
- (c) (5 points) Provide two discount factors $0 < \gamma < 1$ that induce different optimal policies. Explain your answer.



3 Tiger or Ruin?

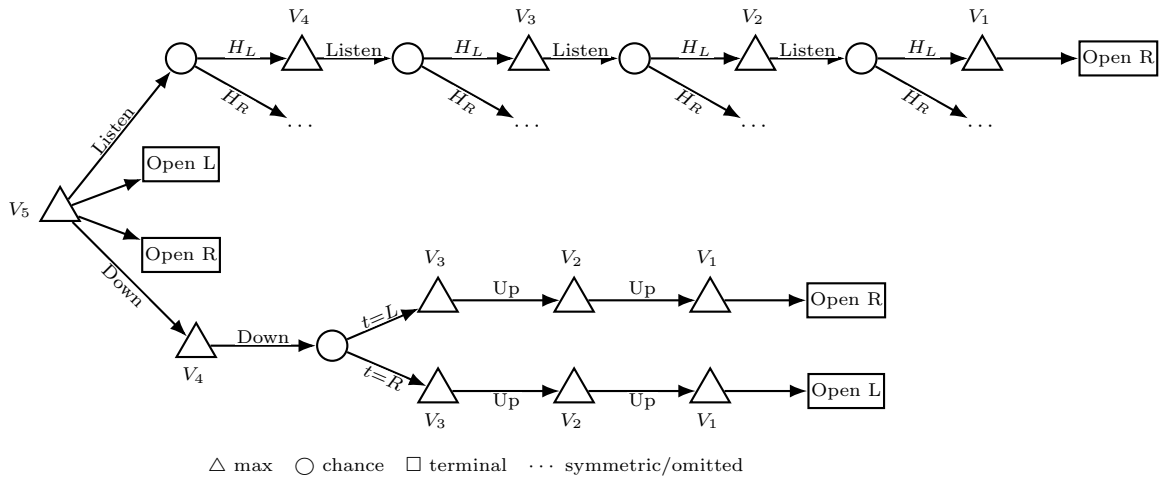
3. Leslie mixed up her notes and somehow ended up with a combination of the tiger problem and the fortune-or-ruin problem. The situation is:

- There are 3 locations in a hallway: 1, 2, 3
- In location 1, you can open the door on the left or the door on the right.
- There is a tiger behind one of the doors and a fortune behind the other.
- If a tiger is behind the door you open you get -100, otherwise you get +10. In either case, the game ends (you get 0 reward for the infinite future).
- You may stand at location 1 and listen. If you do this and detect a tiger on either the left or the right, then the chance that the tiger truly is on the side where you heard it is 0.85. Listening has reward -1.
- You can also move to a neighboring location for a reward of -1.
- In location 3, you will get a perfect observation of where the tiger is.
- In the following, we will only consider finite-horizon values so you can assume $\gamma = 1$.



(a) (4 points) What is the belief space of this problem?

(b) (8 points) Here is a partial horizon 5 expectimax search tree for the combined problem.



Assume that, in b_0 , the agent is in location 1 and believes that tiger is on the left with probability 0.5.

What is the horizon 5 value of selecting *Down* as your first action in the combined problem, and continuing optimally, $Q^5(b_0, \text{Down})$?

Explain your reasoning and write down a numerical expression (but you don't need to evaluate it.)

- (c) (5 points) Assume the optimal horizon 5 value of being in belief state b_0 in the original tiger problem is $V_{\text{tiger}}^5(b_0)$. What must be true of the value of $Q^5(b_0, \text{Down})$ in order for *Down* to be the optimal first action in the combined problem in belief state b_0 ?

- (d) (5 points) Recall that the optimal horizon 3 action from belief $b = (0.5, 0.5)$ in the original tiger problem is *Listen*. What is the optimal horizon 3 action in the combined problem? Explain why your answer does or does not differ from the answer for the horizon 5 case.

4 Ternary Bandits

4. You find yourself in a casino with two slot machines, each of which has a hidden payoff probability p_m (where m is 1 or 2, indicating the machine) which is an element of $\{0.1, 0.5, 0.9\}$.

Unlike a typical bandit problem: You have to pay \$0.50 each time you play, and you have the option of just not playing anymore—taking the nop action.

Furthermore, you can also pay \$.10 to call a guy named Vinnie, and give him the serial number of the machine you're playing. Vinnie will tell you the payoff probability of that machine (that is, 0.1, 0.5, or 0.9) but will only give you the correct answer with probability 0.99. Also, Vinnie has no memory, so each time you call, it's an independent chance the answer is correct.

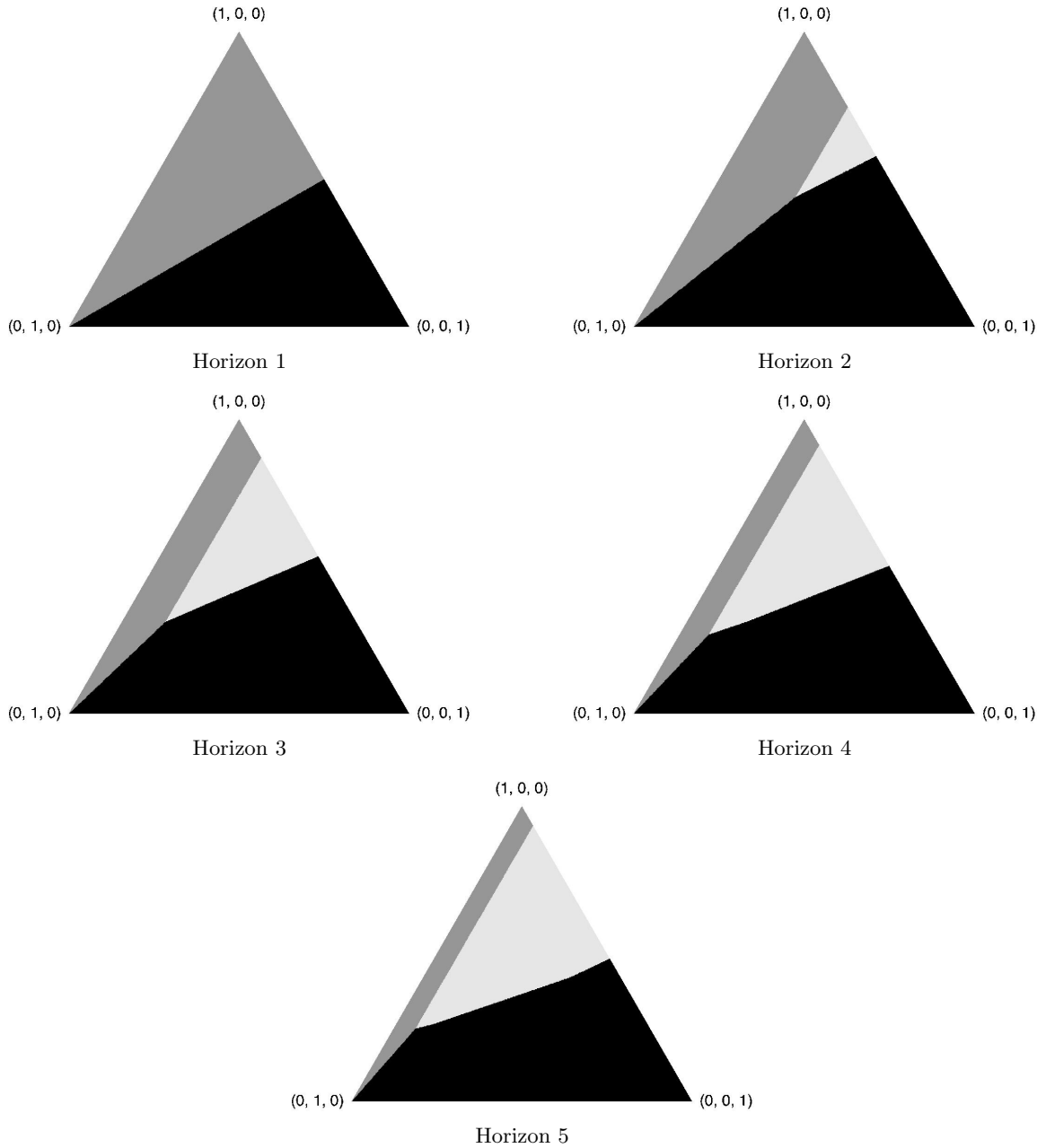
Let's focus on **a single machine**, to start with.

- (a) (3 points) If you start with a belief $(0.1, 0.1, 0.8)$ on p_1 , what is your belief about the probability of winning if you play that machine?

6.4110

Quiz 3, April 27, 2026

(b) (6 points) The following figure contains the optimal horizon 1 through 5 policies. The diagrams have three colors: *black*, *dark gray*, and *light gray*, with each corner of the triangle indicating absolute belief in p_1 being equal to 0.1, 0.5 or 0.9.



i. (3 points) We forgot the legend! Which action corresponds to which color?

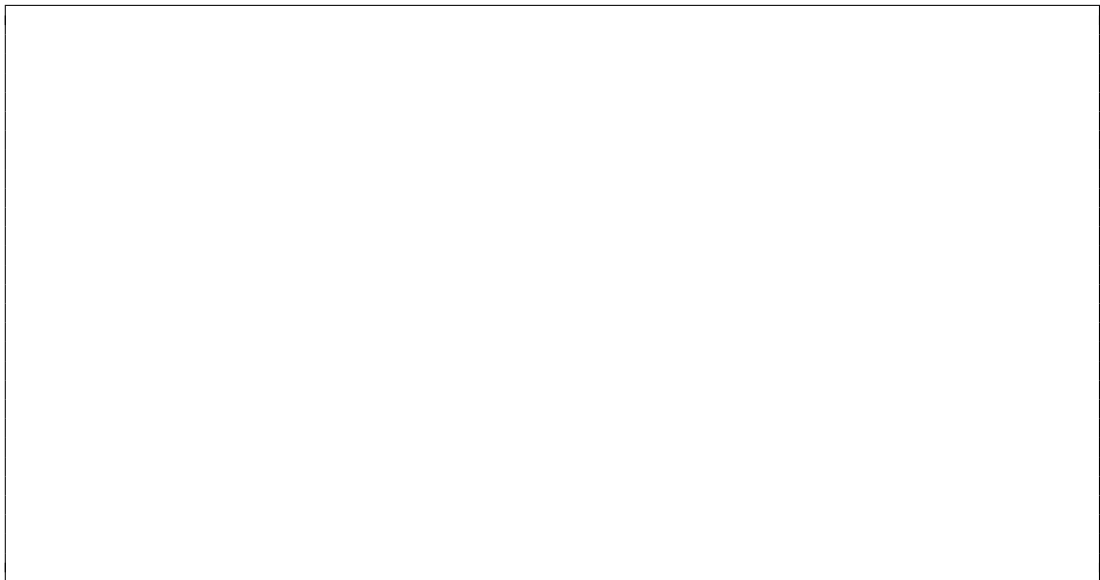
- Dark Gray:** pull arm 1 don't play call vinnie
- Light Gray:** pull arm 1 don't play call vinnie
- Black:** pull arm 1 don't play call vinnie

- ii. (4 points) Why is the black region bigger for horizon 2 than horizon 1?



- iii. (4 points) The stripe along the left side of the triangle is getting thinner as we increase the horizon. Will it ever go completely away? Explain why, in terms of the problem.

Never fully go away Eventually go away



6.4110

Quiz 3, April 27, 2026

Now, you have **two machines**, each with exactly the same set-up as the single one we have studied so far, including the ability to call Vinnie about either machine.

- (c) (6 points) Your friend Sydney thinks that an optimal horizon H strategy in this case would be to compute the optimal horizon H Q functions for both machines, and then to pick the action that has the highest Q value among all 6 choices (three actions on each machine). **Provide an example belief in which this is not true and explain.**

Hint: Consider a scenario where you are fairly certain one machine is good, but highly uncertain about the other.



5 Mini-Project Check

5. In miniproject 4, we were trying to rescue a patient and take them to the hospital in a domain with fast-moving fire. We assumed that the state of the fire was completely observable, but it spread in the domain using random transition dynamics.

For each of the techniques below, **indicate whether we used it** in MP4. If we did, **explain why it was a reasonable choice** for the problem. If we didn't, **explain why it would have been a poor choice**.

- (a) (3 points) Value iteration
 We did We didn't

- (b) (3 points) Most likely observation
 We did We didn't

- (c) (3 points) Monte-Carlo tree search
 We did We didn't