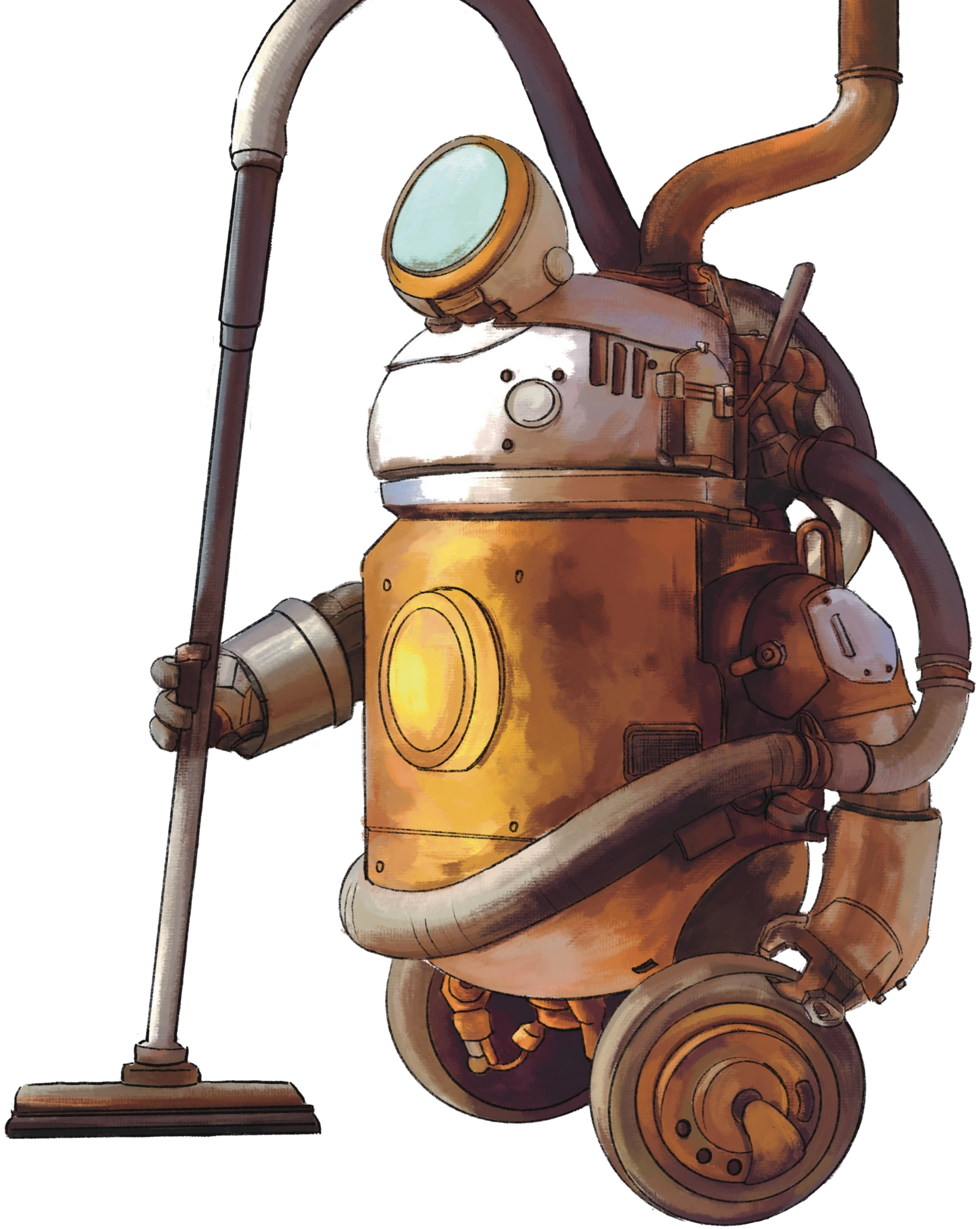
CS 3630, Fall 2025

Lecture 6:

A Vacuum Cleaning Robot:

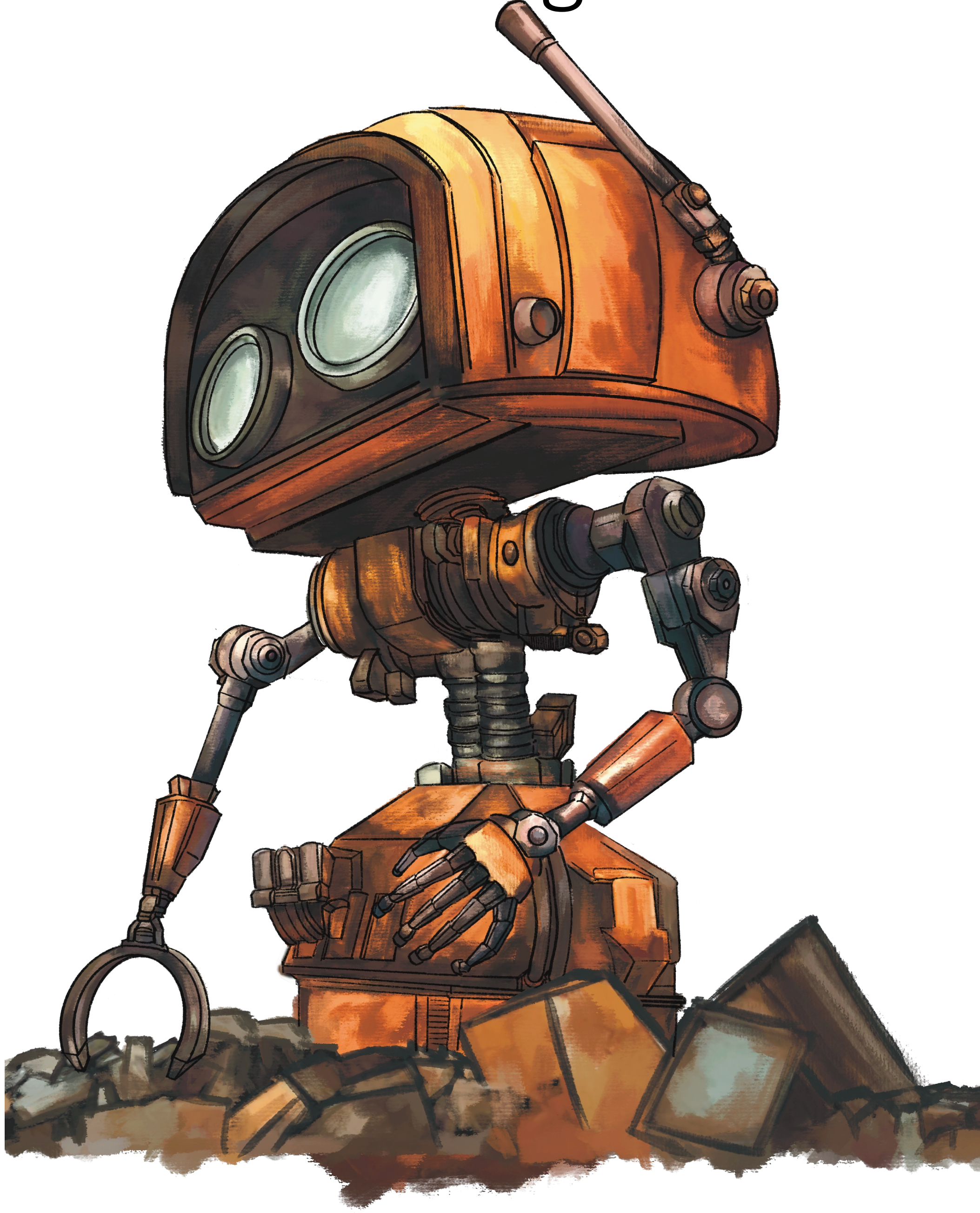
States and Actions



Lectures 3-5 Recap

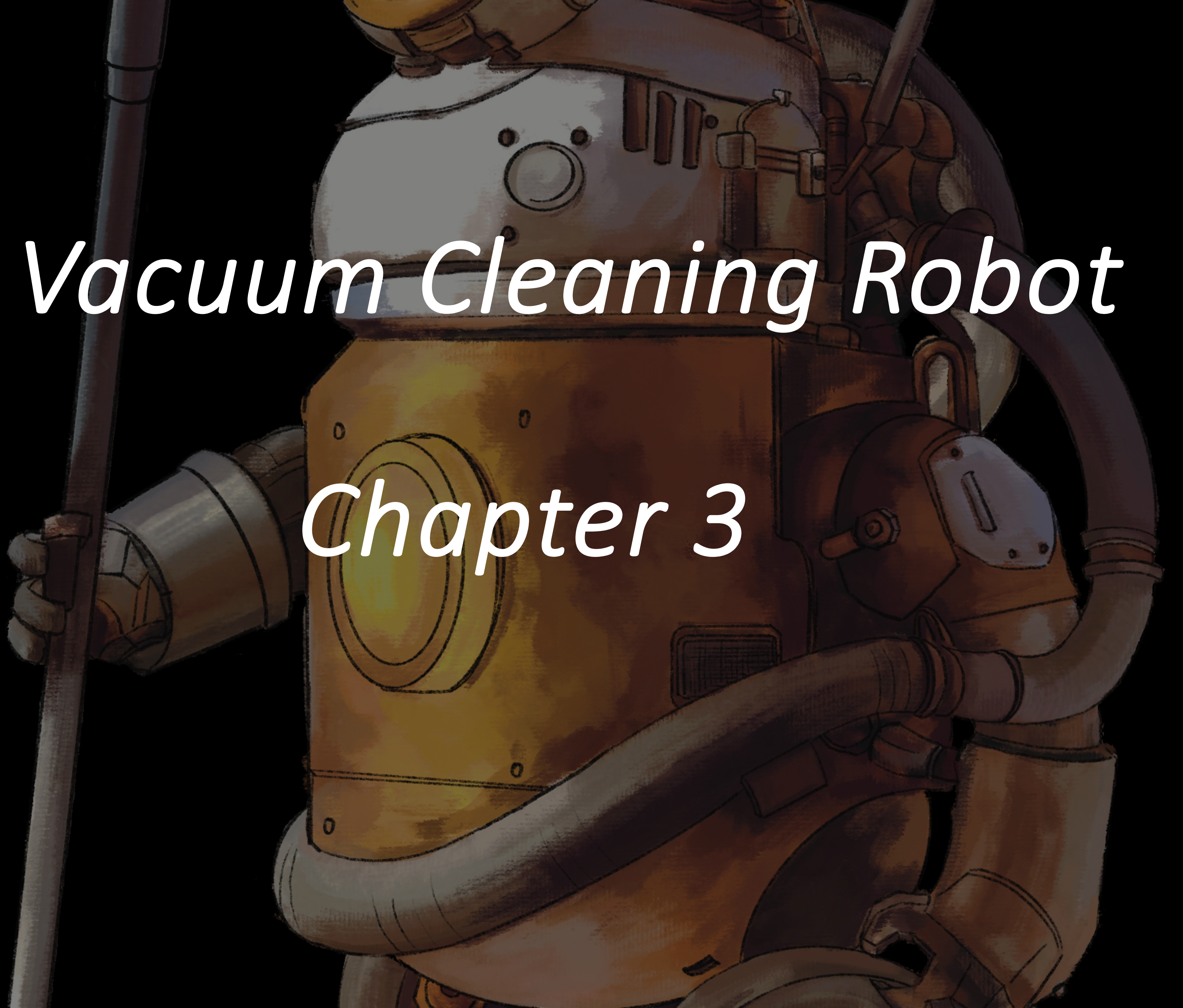
Trash Sorting Robot

- **States** are uncertain:
    - Prior probability distribution on states
- No dependence on past **actions**
- Three simple **sensors**
    - Discrete sensors, discrete conditional distributions
    - Continuous sensor, conditional distribution is Gaussian
- **Perception** using Bayes equation:
    - Bayes inversion equation to infer state from sensors
    - Maximum likelihood estimation (MLE)
    - Maximum a posteriori estimate (MAP)
    - Sensors are conditionally independent, given state
- Deterministic actions make **planning** easy:
    - Formulate decision making as an optimization problem
    - Minimize expected cost, minimize worst-case cost, etc.
- **Learning** prior distributions and sensor models
    - Counting outcomes and using proportions (discrete)
    - Parameter estimation for Gaussian distributions



A Vacuum Cleaning Robot

Chapter 3



Overview

- The **state space** is more interesting, but still discrete.
- **Actions** are not deterministic.
  - Uncertainty in the effects of actions.
  - Probability associated to an action’s effects depends on the current state.
- Very **simple sensing** system
- **Perception** includes both sensing and context
  - An individual measurement from a simple sensor doesn’t provide much information.
  - *History* of sensor observations affects current belief about the world.
- **Planning** is more complex in this scenario.
  - Because effects of actions depend on state, we need to think about more than one action, and about how the effects of actions propagate through time.
  - Because there is uncertainty, we plan to maximize expected reward, not deterministic outcomes or goals.
- Reinforcement **Learning** (RL) is appropriate when we don’t have access to large data sets, and when the robot operates in the same setting for a long period of time.

States

In this chapter, we consider the following scenario:

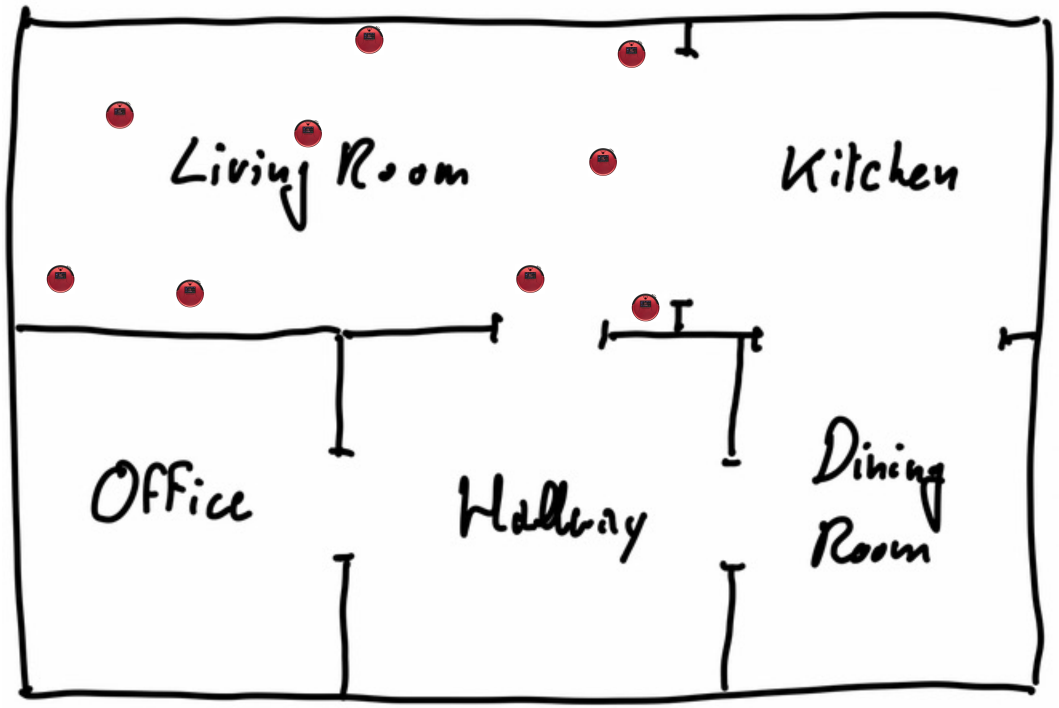
- The robot can move in any direction, so its orientation doesn't matter.
- The robot is equipped with navigation software (which is not perfect), so we won't worry about path planning from room to room.
- To clean a specific room, the robot can execute a preprogrammed motion (maybe boustrophedon, maybe random), so we don't need to worry about the exact position of the robot in a specific room.
- The robot has built-in collision avoidance, so no need to have a detailed map of object locations

➢ ***The room in which the robot is currently located is the only interesting piece of information for this robot.***

State Space

For this robot, the state, X, is defined as the room in which the robot is currently located:

$$X \in \{living\ room, kitchen, office, hallway, dining\ room\}$$



A typical vacuum cleaning robot.

For all of the robot locations shown here, we have:

$$\boldsymbol{X = living\ room}$$

The exact location within the living room is not relevant for this robot.

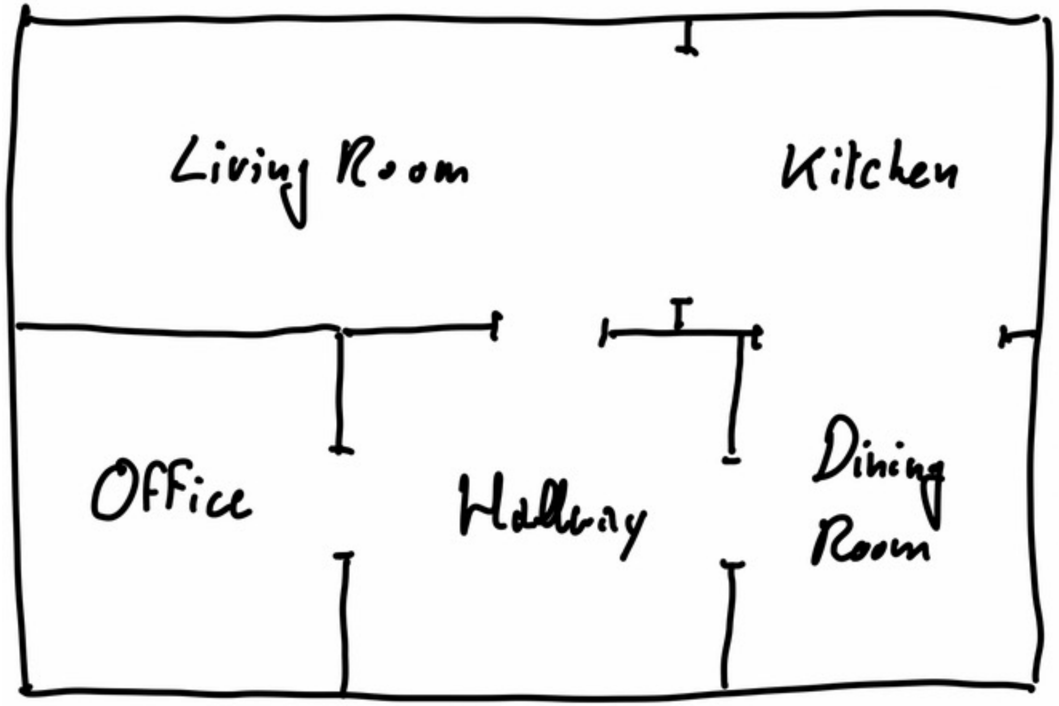
To simplify notation, we'll sometimes write $\boldsymbol{X \in \{L, K, O, H, D\}}$.

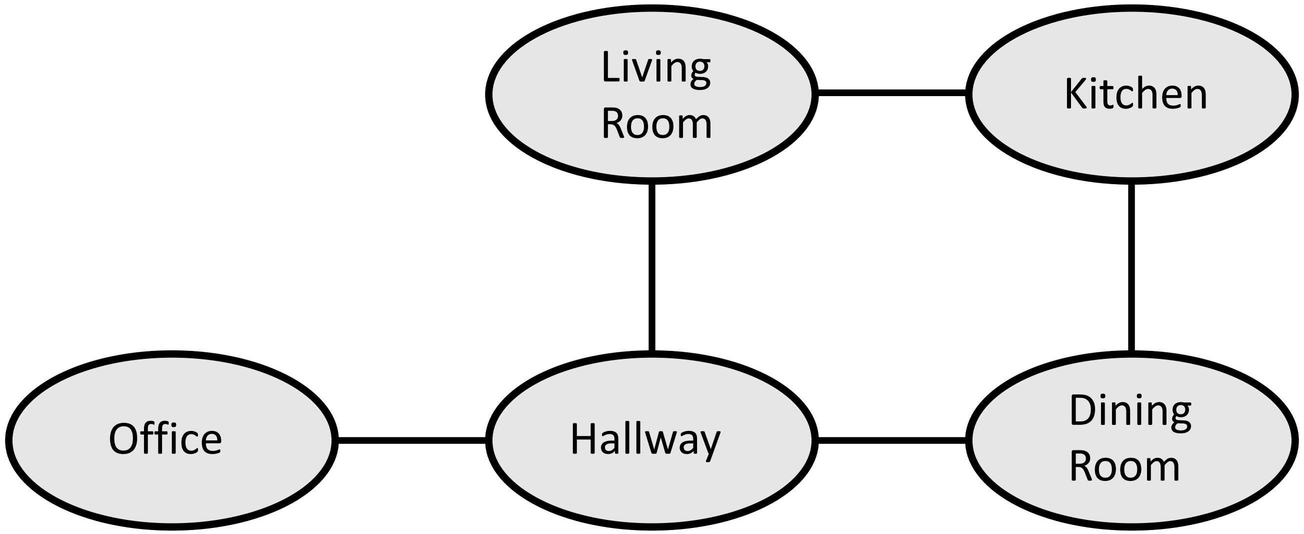
State Space

The state space is the set of all states, along with connectivity information (i.e., neighborhood relationships between states).

In this case, we can represent the state space by a simple undirected graph.

- Our robot can move directly from the *living room* to the *kitchen* or *hallway*, but cannot move directly from the *kitchen* to the *office*.
- This representation will be useful for both planning and for perception.





Prior probability distribution

- For our trash sorting robot, the prior probability distribution on state described our belief (before making any sensor measurements) about the type of object in the work cell.
- For our vacuum cleaning robot, the prior probability distribution on state describes our belief about the ***initial location*** of the robot: *When the robot "wakes up," where will it be?*
- In this chapter, we'll assume that the robot always returns to its charging station in the office after operation (perhaps with the help of a human, or a very smart dog).
- Therefore, our prior distribution on state at the start of the day is given by:

State, *x*	***P(X=x)***
Living room	0
Kitchen	0
Office	1
Hallway	0
Dining room	0

- $P(X = office) = 1$ implies that there is ***no uncertainty*** in our initial state.
- BUT… because there will be uncertainty associated to the effects of actions, this certainty will not long endure after the robot begins its daily activities.

Discrete time systems

- For our trash sorting robot, there was no need to consider the passing of time.
  - Past actions did not affect future performance,
  - Actions were executed in a single time step.
  - The state, X, denoted the state at the present time, and we never needed to represent the state at any other time (neither past nor present).
- For our vacuum cleaning robot, the passing of time is important.
  - We know the location of the robot at the start of the day, but after the robot executes its first actions, there will be uncertainty in the robot's state.
  - The state could change each time the robot executes an action.
  - Sensor measurements depend on state, and state depends on actions; therefore, the sequence in which sensor measurements occur will give us information about the world that can be used for perception.
- Most of the time, nothing interesting happens.
  - We don't need to keep track of the state for all $t \in \mathbb{R}_{\geq 0}$.
  - We only need to keep track of state at discrete time instants, $t \in \{t_0, t_1 \ldots\}$, where $\{t_0, t_1 \ldots\}$ is the set of times at which something "interesting" occurs.
- **We will represent the state at time $\boldsymbol{t}$ by $\boldsymbol{X_t}$, and we'll simplify notation by simply using $\boldsymbol{t \in \{0, 1, 2 \ldots\}}$.**
- The initial state of the robot (i.e., when it wakes up in the morning) is therefore: $X_0 = office$.

Belief state

- It will sometimes be convenient to refer to the entire probability distribution at time t.
- We refer to this distribution as the belief state at time t, denoted by b_t.
- The belief state is a ***row vector*** whose elements correspond to the possible states.
- In our case, there are five possible states, so b_t has five elements.
- At $t = 0$, the belief state is merely our initial distribution:

$$b_0 = [P(X_0 = L), \quad P(X_0 = K) \quad P(X_0 = O) \quad P(X_0 = H) \quad P(X_0 = D)]$$

$$= [0 \quad 0 \quad 1 \quad 0 \quad 0]$$

- The belief state b_{t+1} is conditioned on the initial state x_0 and all actions taken until time t.

$$b_{t+1}^T = \begin{bmatrix} P(X_{t+1} = L \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = K \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = O \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = H \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = D \mid a_1 \dots a_t, x_0) \end{bmatrix}$$

We need to learn about actions…

➢ Note that we use b_{t+1}^T to denote the transpose of b_{t+1} (for formatting purposes).

Actions

- Our vacuum cleaning robot has four actions:

➢ Move *left, right, up*, or *down* (relative to the map of the house)

- Effects of actions are probabilistic.
- Effects of actions depend on the current state.

➢ ***Use conditional probabilities to model the effects of actions.***

- For a specific sequence of actions (e.g., *up, right, down, left*), computing probabilities for states in the distant future seems complicated.

➢ ***Happily, thanks to the Markov property, these computations are not so difficult.***

Actions

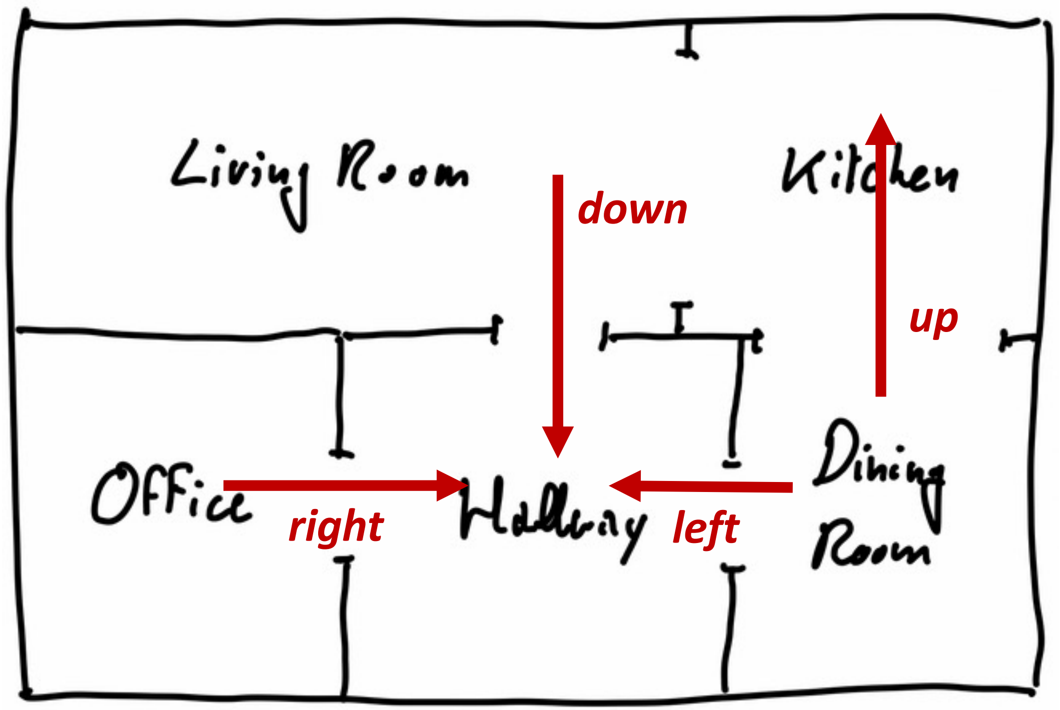
Our robot has four actions:

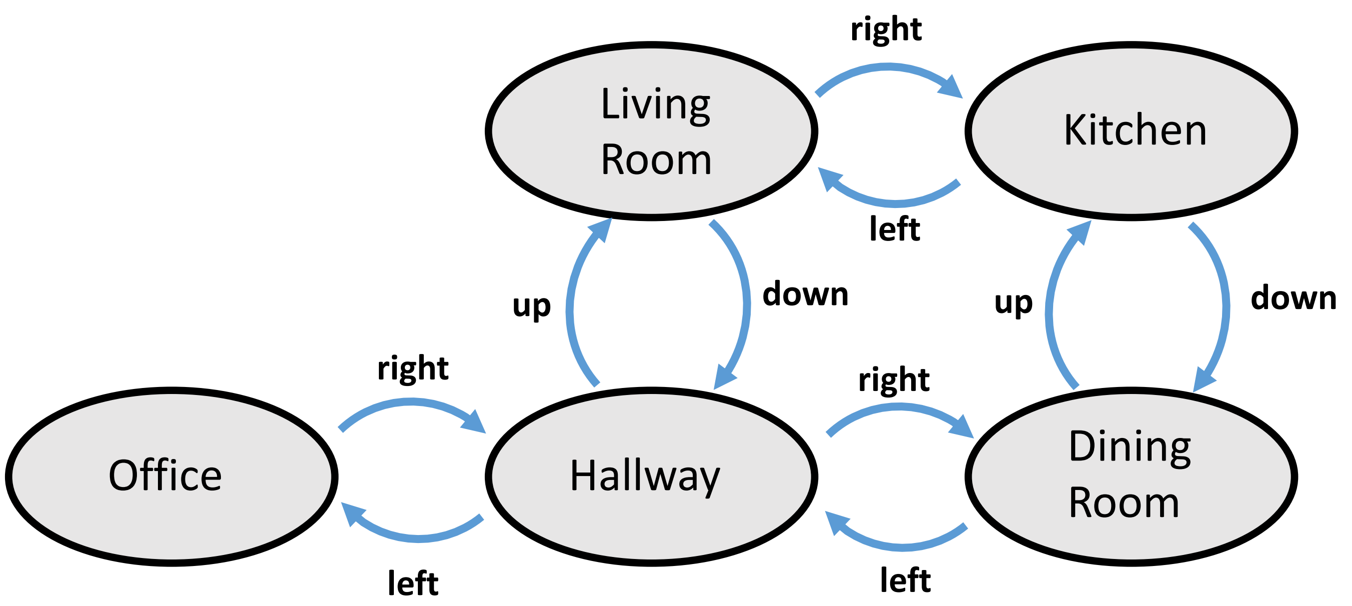
up, down, left, right.

- Effects of actions are context dependent.
- Actions potentially cause a change in state.
- Executing an action in state X_t produces state X_{t+1}

We can represent this by a slight modification to our state space:

- Instead of using an undirected graph, use a directed graph.
- Each edge (u, v) corresponds to an action meant to change the state from $x_t = u$ to $x_{t+1} = v$.
- Sadly, our actions are not deterministic, so we need to do a bit more work.





Uncertainty in the effects of actions

- We will model uncertainty in the effects of actions by using conditional probability distributions.
- In particular, we define the conditional probability distribution for the next state, X_{t+1}, given that the current state, X_t is room x_t, and that action a_t was executed at time t.

$$P(X_{t+1} = x_{t+1} \mid a_t, X_t = x_t)$$

Example: If we are in the *Office* at time *t* and execute the *move right* action, $P(X_{t+1} = H \mid right, X_t = O)$ denotes the conditional probability of arriving to the *Hallway*.

Uncertainty in the effects of actions

- We will model uncertainty in the effects of actions by using conditional probability distributions.
- In particular, we define the conditional probability distribution for the next state, X_{t+1}, given that the current state, X_t is room x_t, and that action a_t was executed at time t.

$$P(X_{t+1} = x_{t+1} \mid a_t, X_t = x_t)$$

Random variable that denotes the state at time $t+1$.

Value taken by X_{t+1} at time $t+1$.

Action taken at time t.

The random state X_t took the value x_t.

Example: If we are in the *Office* at time *t* and execute the *move right* action, $P(X_{t+1} = H \mid right, X_t = O)$ denotes the conditional probability of arriving to the *Hallway*.

The Markov property

- Suppose we start in state $X_0 = O$, execute two actions, *right, up*, and move through the states, $X_1 = H, X_2 = L$.
- What can we say about X_3 if we now execute the action *right*? Or, more formally, what can we say about the conditional probability

$$P(X_3 = x_3 \,|right, up, right, X_0 = O, X_1 = H, X_2 = L)$$

Key Observation:

➢ If we *know* that the robot is in the Living Room at time $t = 2$ and executes the action $a_2 = right$, our belief about X_3 is ***completely independent*** of where the robot may have been at times $t = 0,1$ or of the actions taken at times $t = 0,1$.

➢ *More generally, if we know the current room (aka , X_t) then the history of how the robot came to be in that room will not affect our belief about what happens when the robot executes its next action.*

➢ ***This is an example of a Markov property.***

The Markov property

- Using this Markov property, we can write

$$P(X_3 = x_3 \mid right, up, right, X_0 = O, X_1 = H, X_2 = L) = P(X_3 = x_3 \mid right, X_2 = L)$$

The Markov property

- Using this Markov property, we can write

$$P(X_3 = x_3 \mid right, up, right, X_0 = O, X_1 = H, X_2 = L) = P(X_3 = x_3 \mid right, X_2 = L)$$

What the robot has done before time t.

Where the robot has been before time t.

What the robot does now, at time t.

Where the robot is now, at time t.

Our Markov assumption:

$$\boldsymbol{P(X_{t+1} = x_{t+1} \mid a_0, \dots a_t, X_0 = x_0, \dots, X_t = x_t) = P(X_{t+1} = x_{t+1} \mid a_t, X_t = x_t)}$$

Next Lecture: More Vacuum Cleaning Robot Stuff

- Uncertainty in actions: Markov Decision Process (MDP)
- Uncertainty in sensing for a sequence of measurements: Hidden Markov Model (HMM)
- Planning using Value Iteration
- Reinforcement Learning (RL)