AGENTS

Lara J. Martin (she/they)
TA: Aydin Ayanzadeh (he)

9/5/2023

CMSC 671

By the end of class today, you will be able to:

1. Explain whether an agent is acting rationally or not
2. Describe environments using the language of the field
3. Categorize agents based on their capabilities and behavior

Most slides by Dr. Cassandra Kent & Dr. Mark Riedl

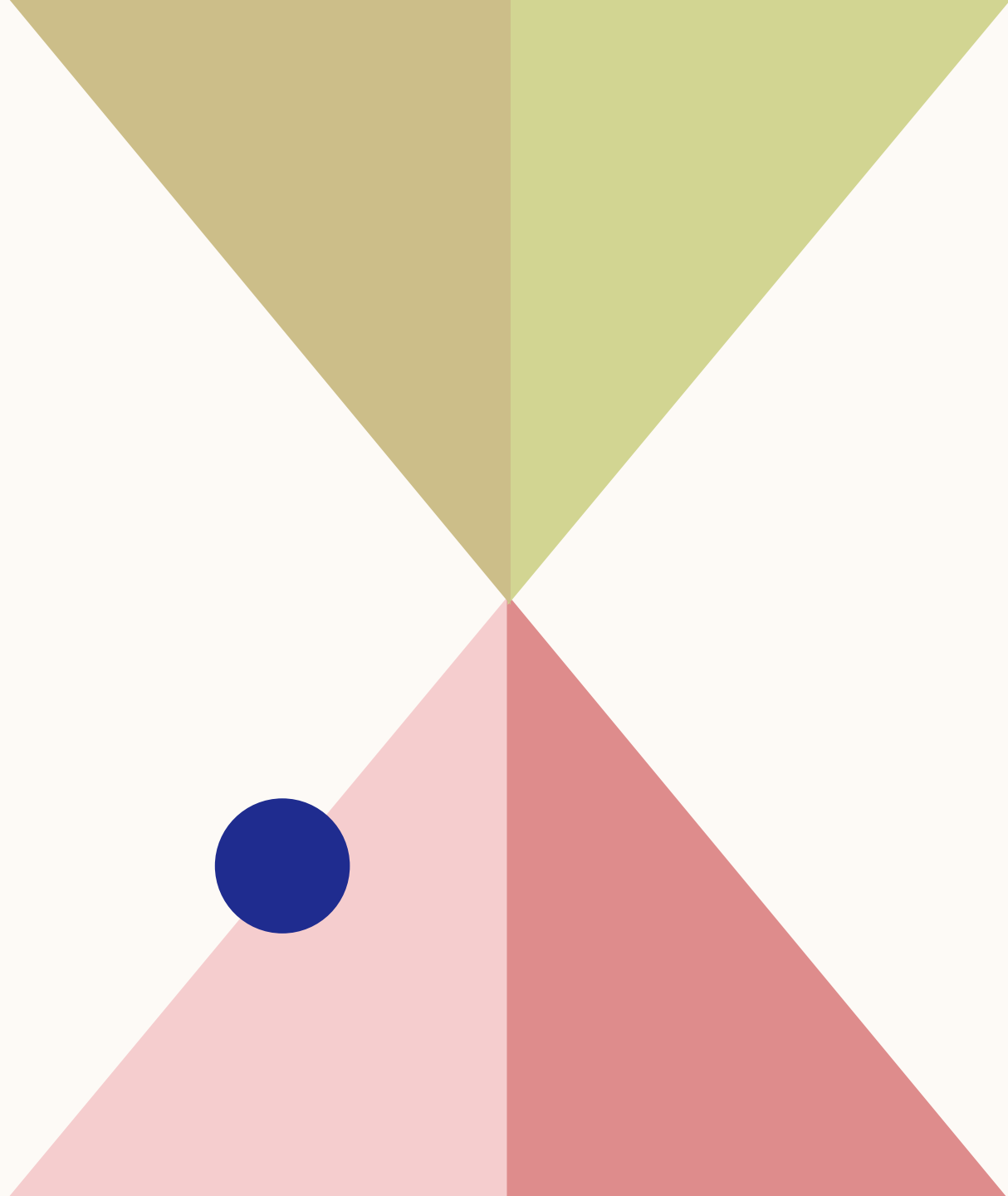
SCHEDULE

Logistics

Recap

Types of environments

Types of agents



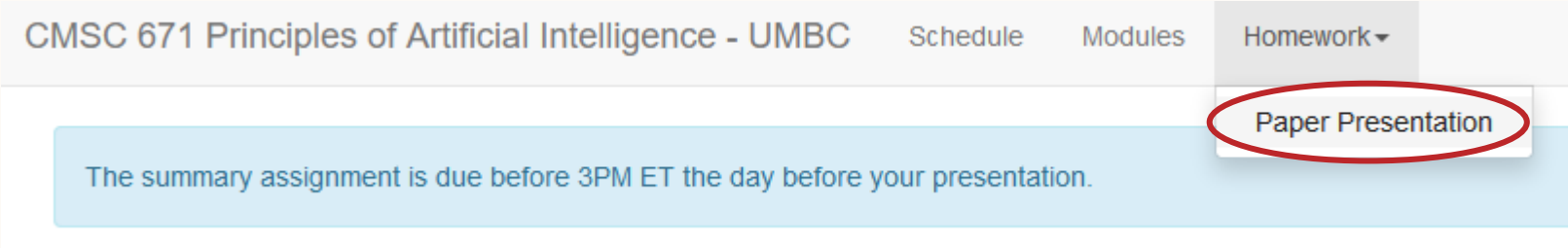
GRADE WEIGHTS CHANGED

Grading

- 5% Class Knowledge Checks
- 10% Paper Presentation
- 40% Homeworks (4 in total; 10% each)
- 20% Midterm
- 25% Final Project

PAPER PRESENTATIONS

https://laramartin.net/Principles-of-AI/homeworks/presentations.html



The summary assignment is due before 3PM ET the day before your presentation.

Paper Presentation (10%)

This assignment is to show you the modern uses of older AI methods and give you an entry point for how to critically read an academic paper.

Learning Objectives

In this assignment, you will

- find reputible research articles from a specific AI area that you find interesting
- recognize & synthesize key points of a research paper
- communicate key findings from a research paper

Instructions

- Pick a Module from 1-4: sign-up sheet
  - Modules will be assigned to be as evenly distributed across the class as possible, keeping in mind your preferences.
  - Once you are assigned a Module, you will be told the approximate date when your presentation will be. (Since the lecture material moves around as the course progresses, the presentation dates might move as well.)
- [1 pt] **Find a recent paper** (published within the past 5 years) from a reputable AI conference or journal. Your selection should be submitted to Lara & Aydin a week before your presentation & summary are due so that we can verify that your choice is a peer-reviewed article that is relevant to the Module. When you submit your selection, you will provide 1) the name of the article, 2) the authors of the article, 3) the link where the article can be found online, and 4) what class topic (i.e., lesson title) the article is relevant to.
- [5 pts] **Summarize** the paper in a 1-page report. Please include, *in your own words*:
  - what are the main findings of the paper?
  - how does the paper relate to the class?
  - what are the strengths of the paper?
  - what are the weaknesses of the paper?
  - are there any ethical concerns that people should consider if they were to replicate the paper or use any of the methods/data/etc. that are introduced in the paper?
- [1 pt] **Present what you learned from the paper** to small groups in class. You're welcome to create a few slides to help you present to your group, but this is not required. We will take abou 10 minutes talking in separate groups and then 5 minutes coming together to share with the whole class.
- 7 points total

Generative AI Policy

If you use ChatGPT (or similar chatbots or AI-based generation tools), you must describe exactly how you used it, including providing the prompt, the original generation, and your edits. This applies to prose, code, or any form of content creation. Not disclosing is an academic integrity violation. If you do disclose, your answer may receive anywhere from 0 to full credit, depending on the extent of substantive edits, achievement of the learning objectives, and overall circumvention of those objectives.

Due this Friday!

Due a week before presentation

Due the day before presentation

Presentation day

RECAP: KNOWN ISSUES OF GPT

- Bad reproducibility
- Copyright issues
- Can’t explain what it’s doing
- Can’t remember things long term
- Confident bullshitter

Based on a slide by Dr. Frank Ferraro

RECAP: WHY AI?

Engineering

- To get machines to do a wider variety of useful things
  - Understand spoken natural language
  - Recognize individual people in visual scenes
  - Find the best travel plan for your vacation

Cognitive Science

- Help understand how natural minds work
  - Visual perception, memory, learning, language, etc.

Philosophy

- As a way to explore interesting (and important) philosophical questions

Slide by Dr. Cynthia Matuzek

BUILDING RATIONAL AGENTS

DEFINITION OF AI

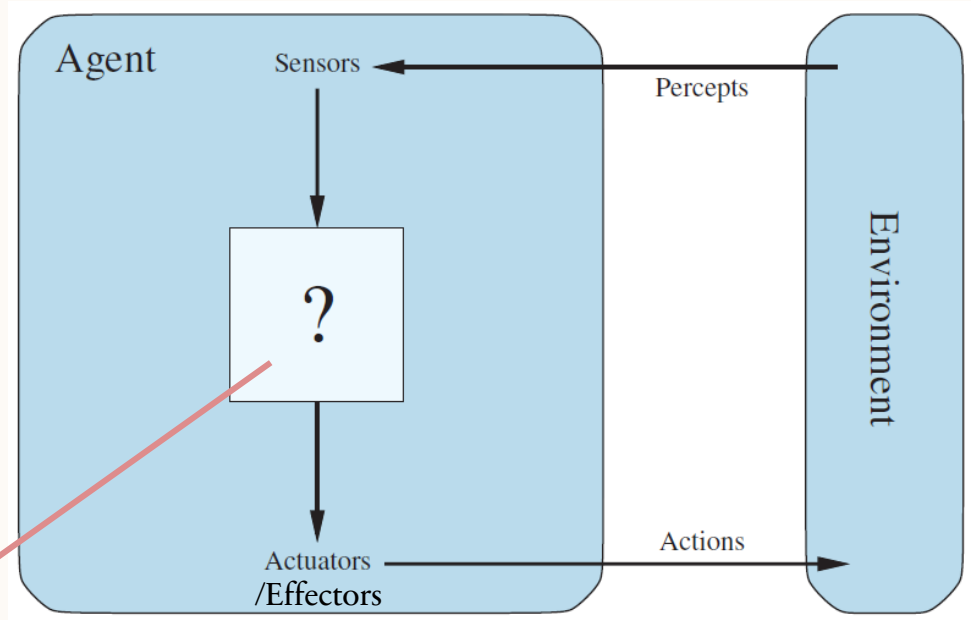
The study of intelligence – both to **understand** and **build** intelligent agents, which can:

- Act humanly
- Think humanly
- Think rationally
- Act rationally

Determine what action to take to achieve the best outcome.

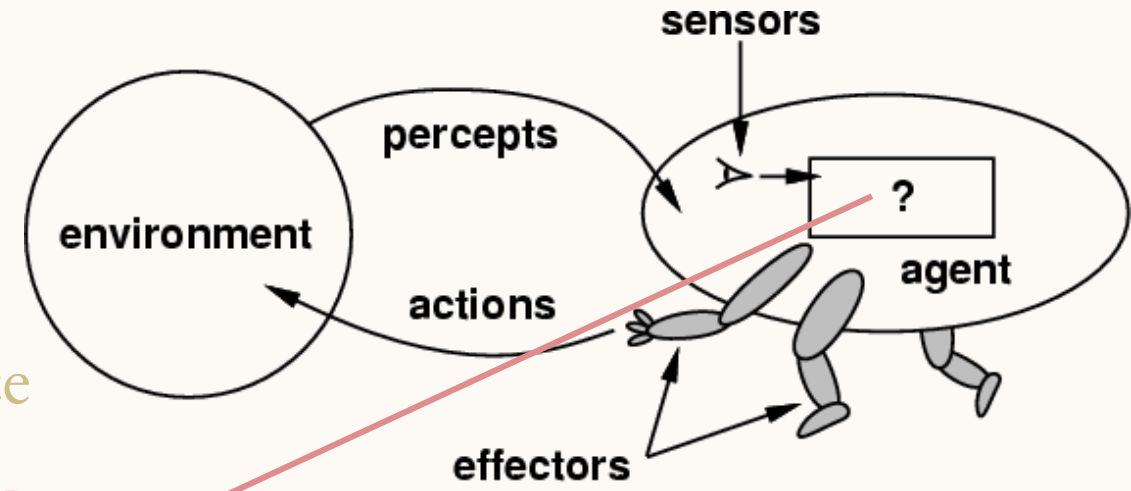
AGENT DEFINITION

- **Agent:** anything that perceives its environment through sensors, and acts on its environment through actuators
- **Percept:** input at an instant
- **Percept sequence:** history of inputs
- **Agent function:** mapping of percept sequence to action
- **Agent program:** (concise) implementation of an agent function



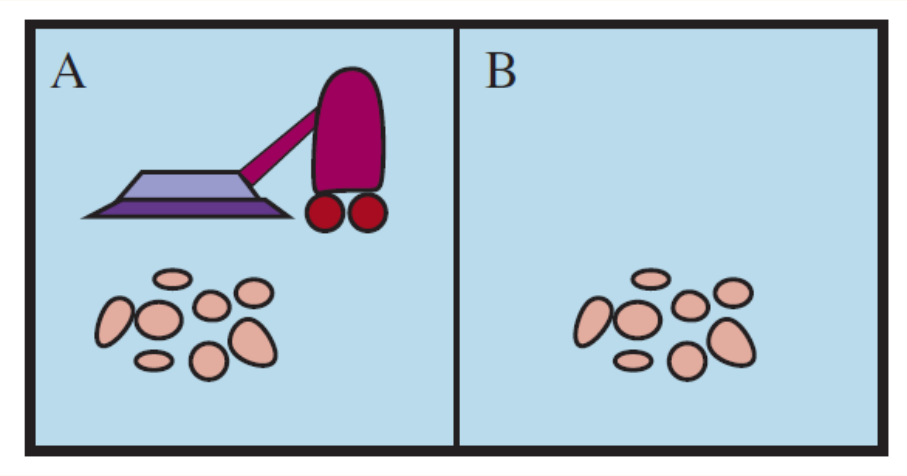
AGENT DEFINITION

- **Agent:** anything that perceives its environment through sensors, and acts on its environment through actuators
- **Percept:** input at an instant
- **Percept sequence:** history of inputs
- **Agent function:** mapping of percept sequence to action
- **Agent program:** (concise) implementation of an agent function



Picture from Dr. Cynthia Matuszek, source unknown

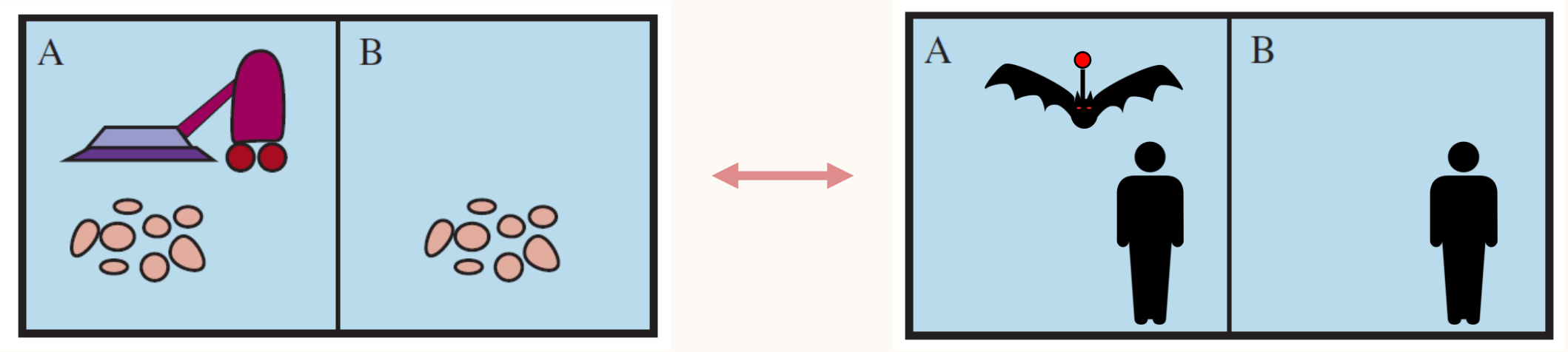
EXAMPLE AGENT: VACUUM-CLEANER WORLD



- Agent can be in 1 of 2 locations (A or B)
- Agent's actions: [*Left*, *Right*, *Suck*, *Wait*]

EXAMPLE AGENT: ~~VACUUM-CLEANER~~ WORLD

Vampire

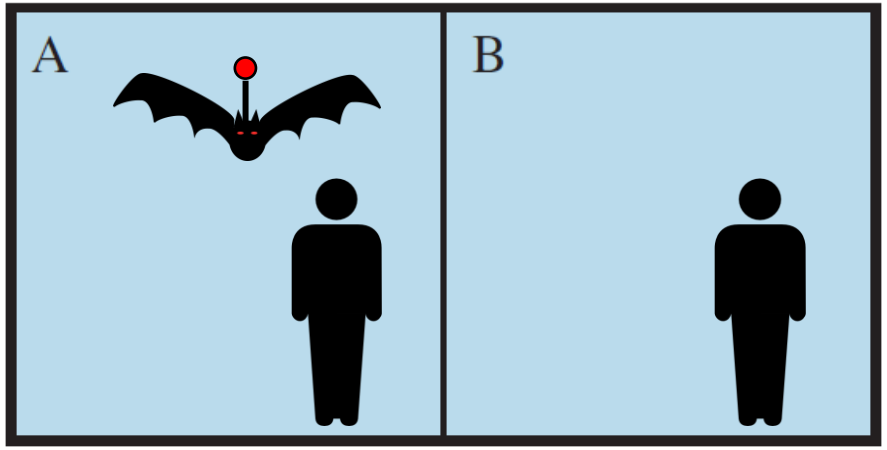


- Agent can be in 1 of 2 locations (A or B)
- Agent's actions: [*Left*, *Right*, *Suck*, *Wait*]

EXAMPLE AGENT: VAMPIRE WORLD

Agent program:

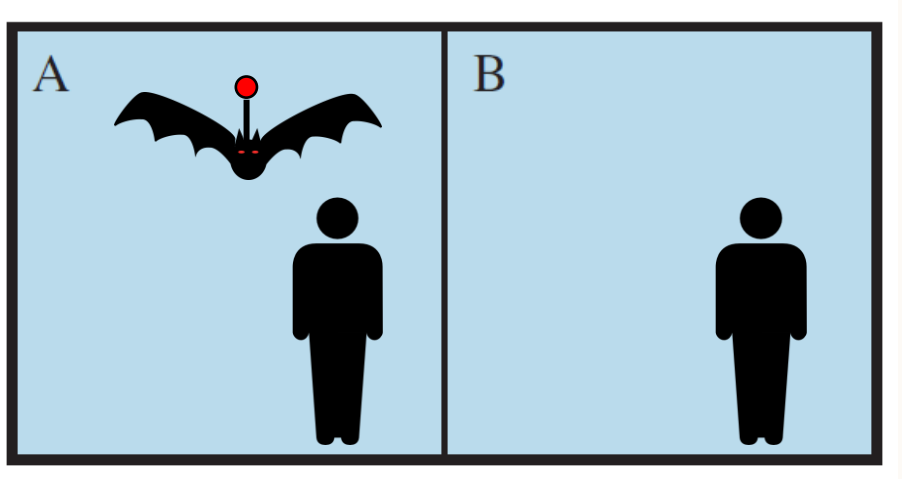
Percept sequence	Action
[A, Empty]	Right
[A, Human]	Suck
[B, Empty]	Left
[B, Human]	Suck
…	…
[A, Empty], [B, Human]	Suck
[A, Empty], [B, Empty]	Left
…	…



EXAMPLE AGENT: VAMPIRE WORLD

Agent program:

```
if [A, Empty]: return Right
if [B, Empty]: return Left
if [A or B, Human]: return Suck
```



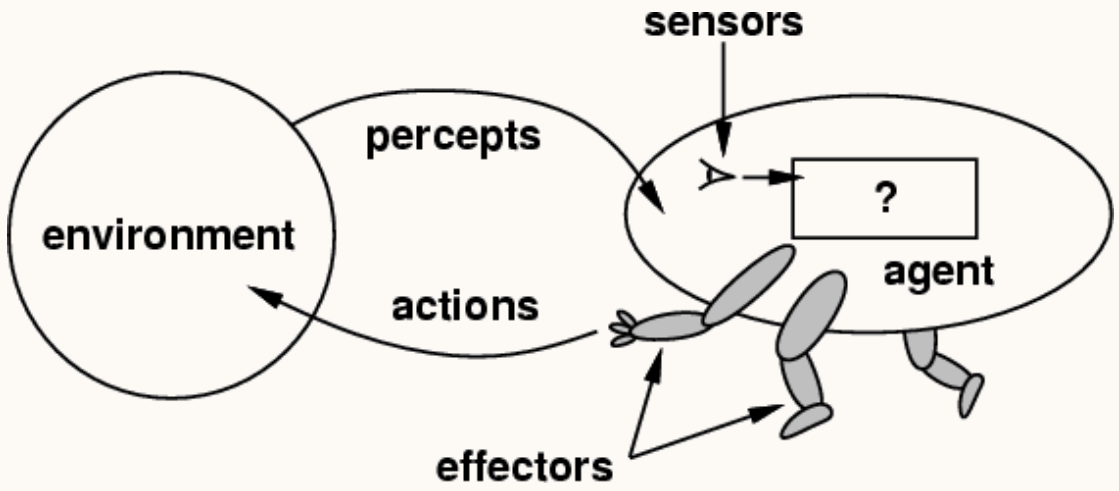
Is this a rational (vampire) agent?

It depends.

WHAT MAKES AN AGENT RATIONAL?

A agent is *rational* when…

it optimizes its actions based on information it gets from its environment (i.e., percepts) in order to **achieve a specific goal**.



DEFINING THE TASK ENVIRONMENT

Agent: Vampire

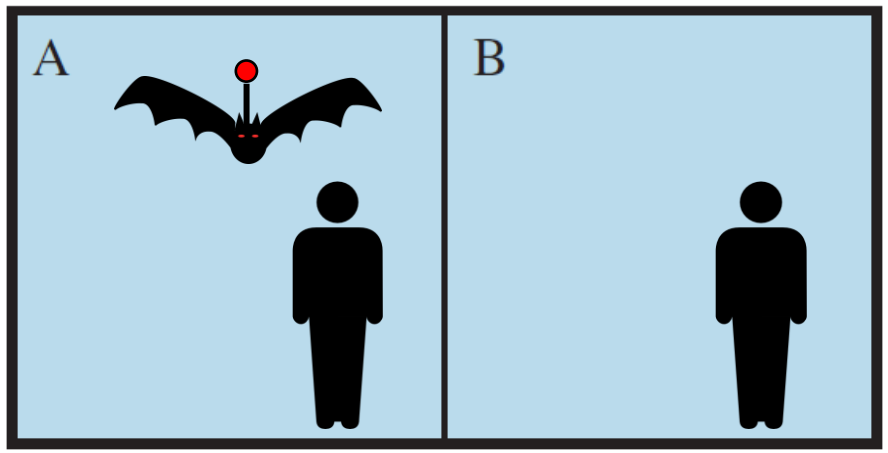
Goal

Performance measure: Suck as much blood over time as possible

Environment: Location, humans

Actuators: Flying, sucking

Sensors: Short-range sonar (human detection)




```
if [A, Empty]: return Right
if [B, Empty]: return Left
if [A or B, Human]: return Suck
```

Now, is this a rational (vampire) agent?

Yes.

DEFINING THE TASK ENVIRONMENT

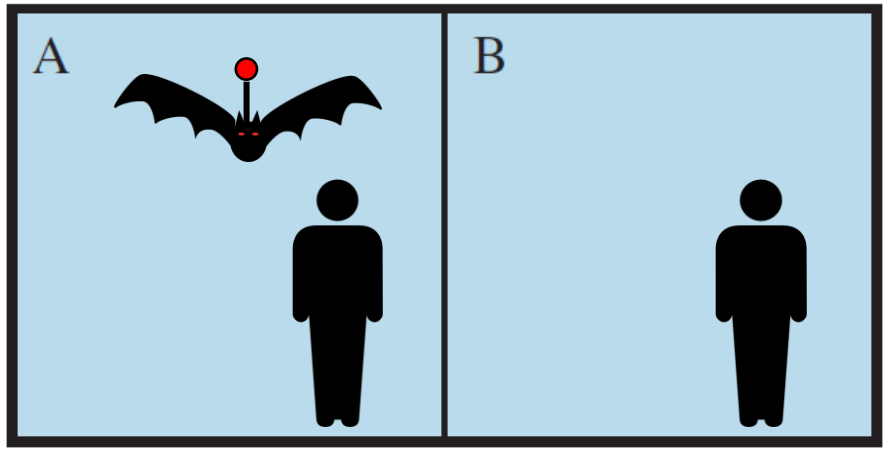
Agent: Vampire

Performance measure: Be an ethical vampire, and avoid humans

Environment: Location, humans

Actuators: Flying, sucking

Sensors: Short-range sonar (human detection)




```
if [A, Empty]: return Right
if [B, Empty]: return Left
if [A or B, Human]: return Suck
```

Now, is *this* a rational (vampire) agent?

No.

DEFINING THE TASK ENVIRONMENT

Agent: Chess Player
Performance measure:

Environment:

Actuators:
Sensors:

Garry Kasparov vs. Deep Blue, 1997

DEFINING THE TASK ENVIRONMENT

Agent: Chess Player

Performance measure: Checkmate the opponent before they checkmate you

Environment: Chess board, pieces, rules of chess

Actuators: Move a piece

Sensors: Current board state

Garry Kasparov vs. Deep Blue, 1997

DEFINING THE TASK ENVIRONMENT

Agent: Search Engine

Performance measure:

Environment:

Actuators:

Sensors:

DEFINING THE TASK ENVIRONMENT

Agent: Search Engine

Performance measure: Return relevant results, with the best results appearing earlier in the list

Environment: Web pages on the Internet

Actuators: Generate page list

Sensors: Read query, read page text, read page metadata

YOUR TURN: DEFINING THE TASK ENVIRONMENT

Agent: Self-driving Car
Performance measure:

Environment:

Actuators:

Sensors:

Tesla's semi-self-driving car, 2020

YOUR TURN: DEFINING THE TASK ENVIRONMENT

Agent: Self-driving Car

Performance measure: Get passenger to destination, safely, legally, comfortably, quickly, …

Environment:

Actuators:

Sensors:

Tesla's semi-self-driving car, 2020

YOUR TURN: DEFINING THE TASK ENVIRONMENT

Agent: Self-driving Car

Performance measure: Get passenger to destination, safely, legally, comfortably, quickly, …

Environment: Roads, vehicles, pedestrians, passengers, …

Actuators:

Sensors:

Tesla's semi-self-driving car, 2020

YOUR TURN: DEFINING THE TASK ENVIRONMENT

Agent: Self-driving Car

Performance measure: Get passenger to destination, safely, legally, comfortably, quickly, …

Environment: Roads, vehicles, pedestrians, passengers, …

Actuators: steering, acceleration, braking, horn, signals, lights, …

Sensors:

Tesla's semi-self-driving car, 2020

YOUR TURN: DEFINING THE TASK ENVIRONMENT

Agent: Self-driving Car

Performance measure: Get passenger to destination, safely, legally, comfortably, quickly, …

Environment: Roads, vehicles, pedestrians, passengers, …

Actuators: steering, acceleration, braking, horn, signals, lights, …

Sensors: cameras, LIDAR, GPS, speedometer, accelerometers, …

Tesla's semi-self-driving car, 2020

WAYS TO DESCRIBE ENVIRONMENTS

Observability

Determinism

Static

Discreteness

Episodic

Number of Agents

DESCRIBING THE ENVIRONMENT: OBSERVABILITY

Fully Observable

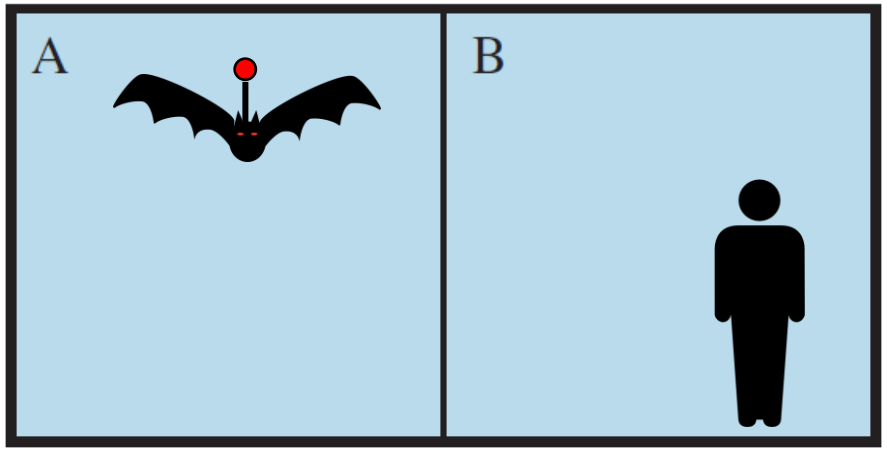
- Can sense everything in the environment without error
- *Examples*: chess, image classification

Partially Observable

- Can only sense some information
- Two causes:
  - Incomplete data
  - Noise (sensor errors)
- *Examples*: poker, vehicle localization

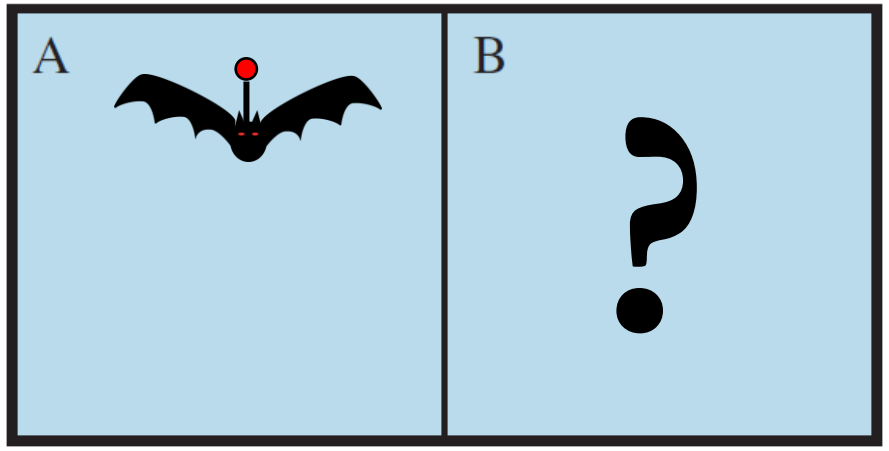
OBSERVABILITY IN VAMPIRE WORLD

Fully Observable



- Agent can observe the complete environment state

Partially Observable



- Agent can only observe part of the environment

DESCRIBING THE ENVIRONMENT: DETERMINISM

Deterministic

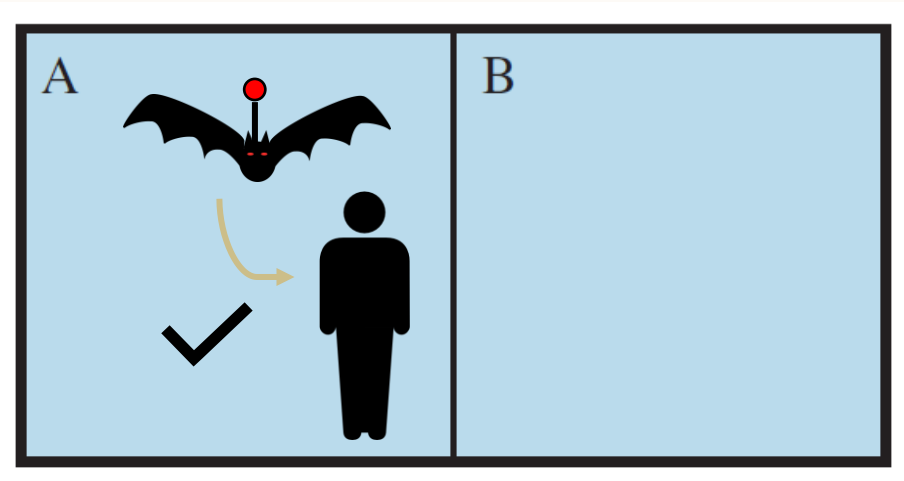
- Environment changes in exactly one way as a result of an agent's actions
- *Examples*: go, StarCraft

Stochastic

- Randomness
- Action uncertainty
- Partial-observability
- *Examples*: rock-paper-scissors, bin picking robot

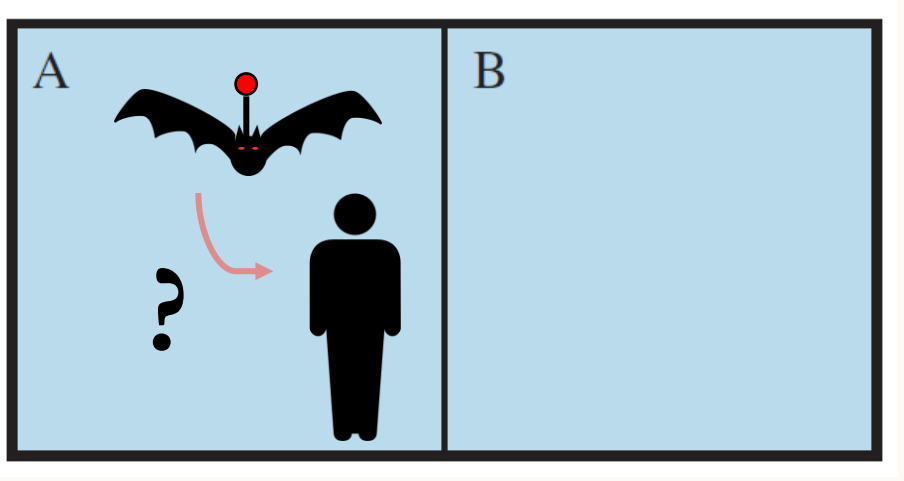
DETERMINISM IN VAMPIRE WORLD

Deterministic



- *Suck* action will always succeed

Stochastic



- *Suck* action may fail, chance of human fighting off vampire agent

DESCRIBING THE ENVIRONMENT: STATIC

Static

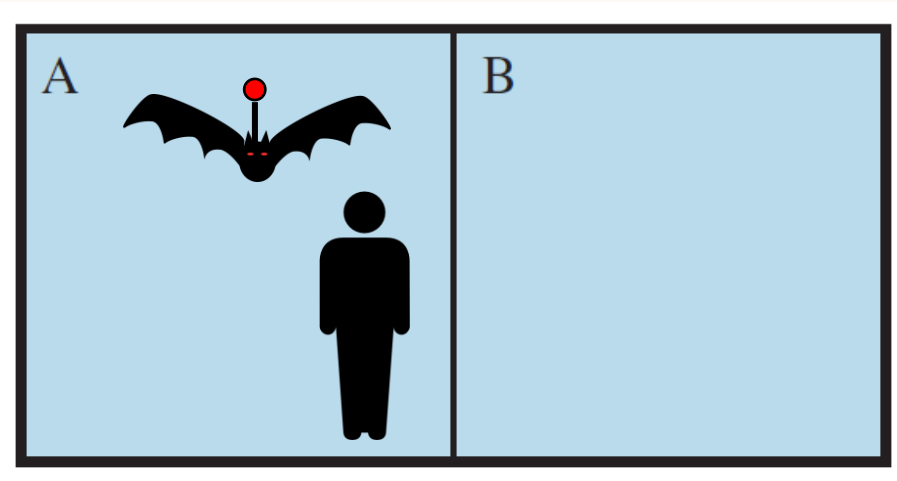
- World will not change while an agent is deliberating
- *Examples*: Sudoku, automatic character recognition

Dynamic

- World can change while an agent is deliberating
- *Examples*: Dota, robot answering machine

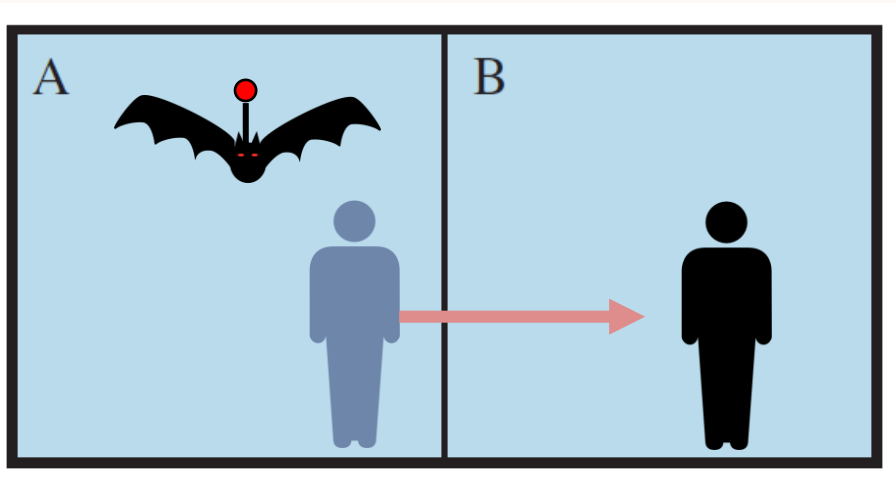
STATIC IN VAMPIRE WORLD

Static



- Humans will wait while vampire agent decides what to do

Dynamic



- Humans move around independently of the decision time step

DESCRIBING THE ENVIRONMENT: DISCRETENESS

Discrete

- World broken up into a finite number of discrete chunks
- *Examples*: image generator, go

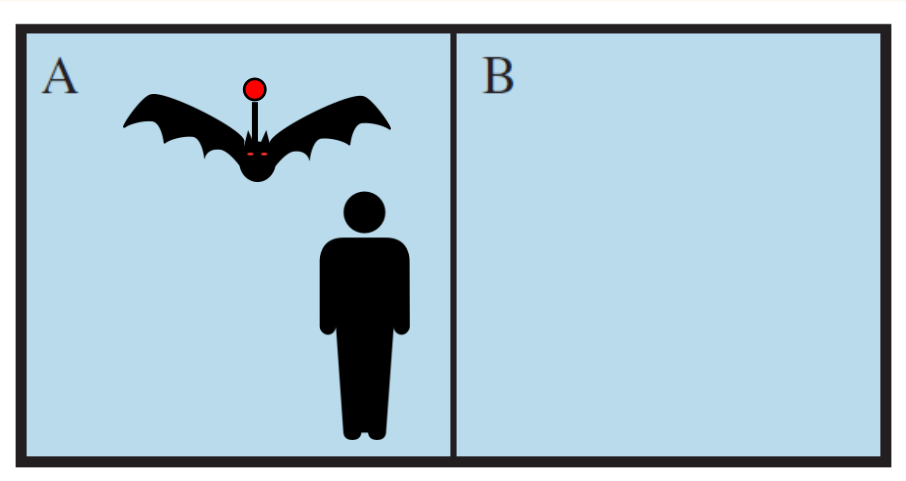
Continuous

- Infinite number of chunks
- Infinite gradations of values
- *Examples*: image labeler, StarCraft

Can describe space, time, actions, and percepts

DISCRETENESS IN VAMPIRE WORLD

Discrete



- Environment divided into grid cells

Continuous

- Environment represented as a continuous space

DESCRIBING THE ENVIRONMENT: EPISODIC

Episodic

- Percept sequence does not affect action selection
- History isn't required
- *Examples*: image classification, test question solver

Sequential

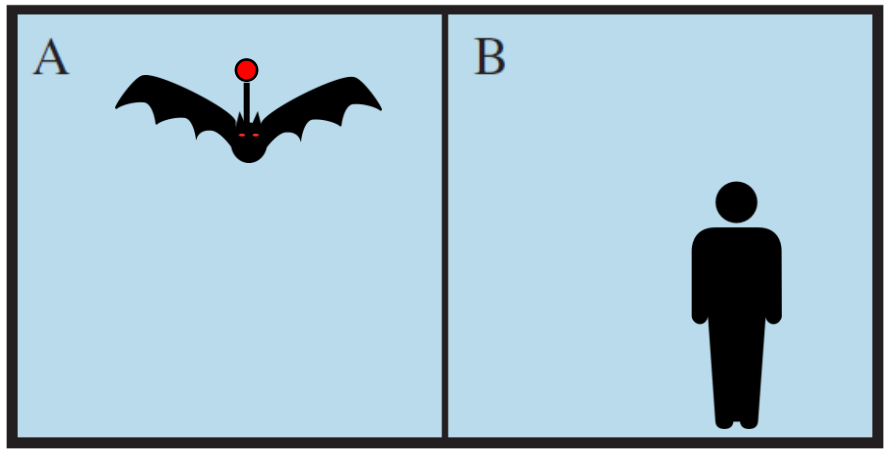
- Percept sequence matters
- Current choice of action will affect future choices
- *Examples*: video tracking, tic-tac-toe

EPISODIC IN VAMPIRE WORLD

Episodic

?

Sequential



- Generally the case for physically situated agents that move through an environment

DESCRIBING THE ENVIRONMENT: AGENTS

Single Agent

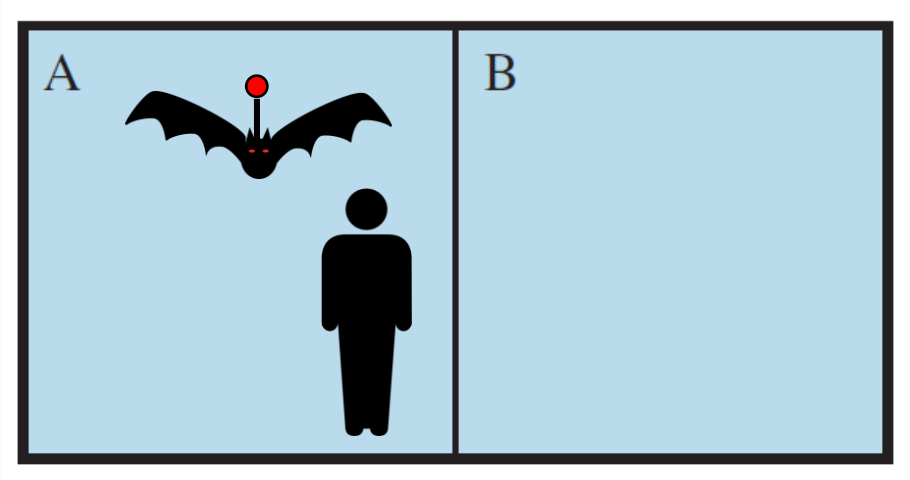
- You are the only agent in the world
- *Examples*: search engine, task scheduler

Multi Agent

- There are other autonomous agents in the world that affect what you're trying to do
- Can be
  - Cooperative
  - Competitive
- *Examples*: chat bot, auction bidder

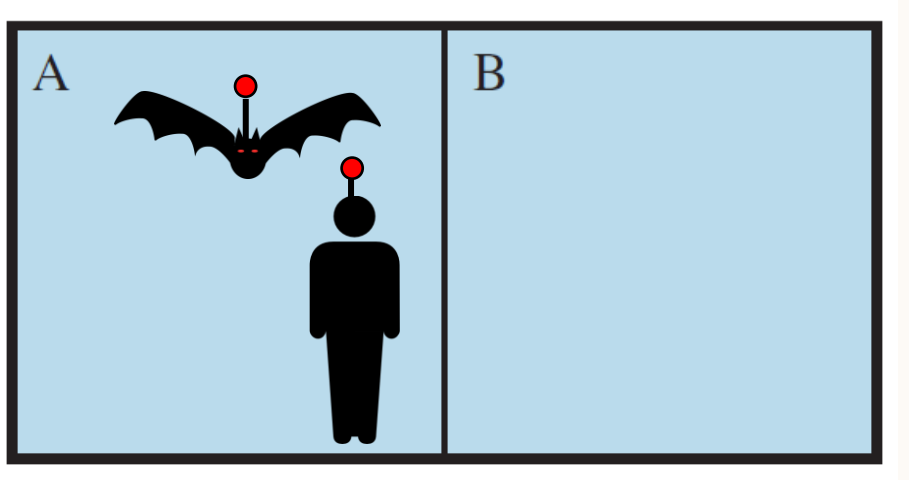
AGENTS IN VAMPIRE WORLD

Single Agent



- Treat human as just another feature of the environment

Multi Agent



- Model human as another agent, with their own autonomy to interact with our agent

DESCRIBING A SELF-DRIVING CAR ENVIRONMENT

- Partially observable
- Stochastic
- Dynamic
- Sequential
- Multi-agent

FOR NEXT CLASS

- Fill out the paper presentation survey
- Read Chapter 3.1-3.4