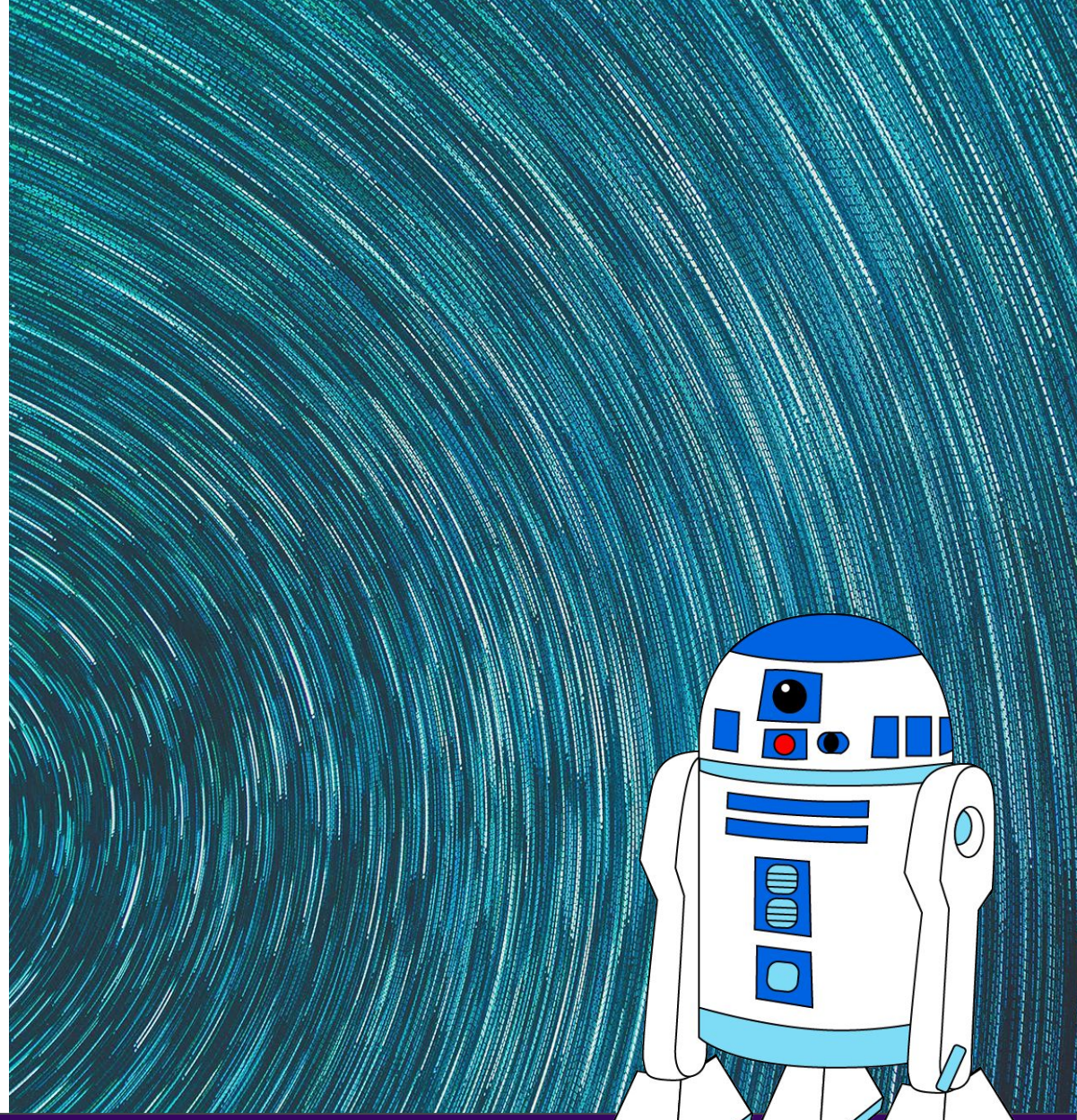
CIS 421/521:
ARTIFICIAL INTELLIGENCE

Rational Agents



Penn Engineering
University of Pennsylvania

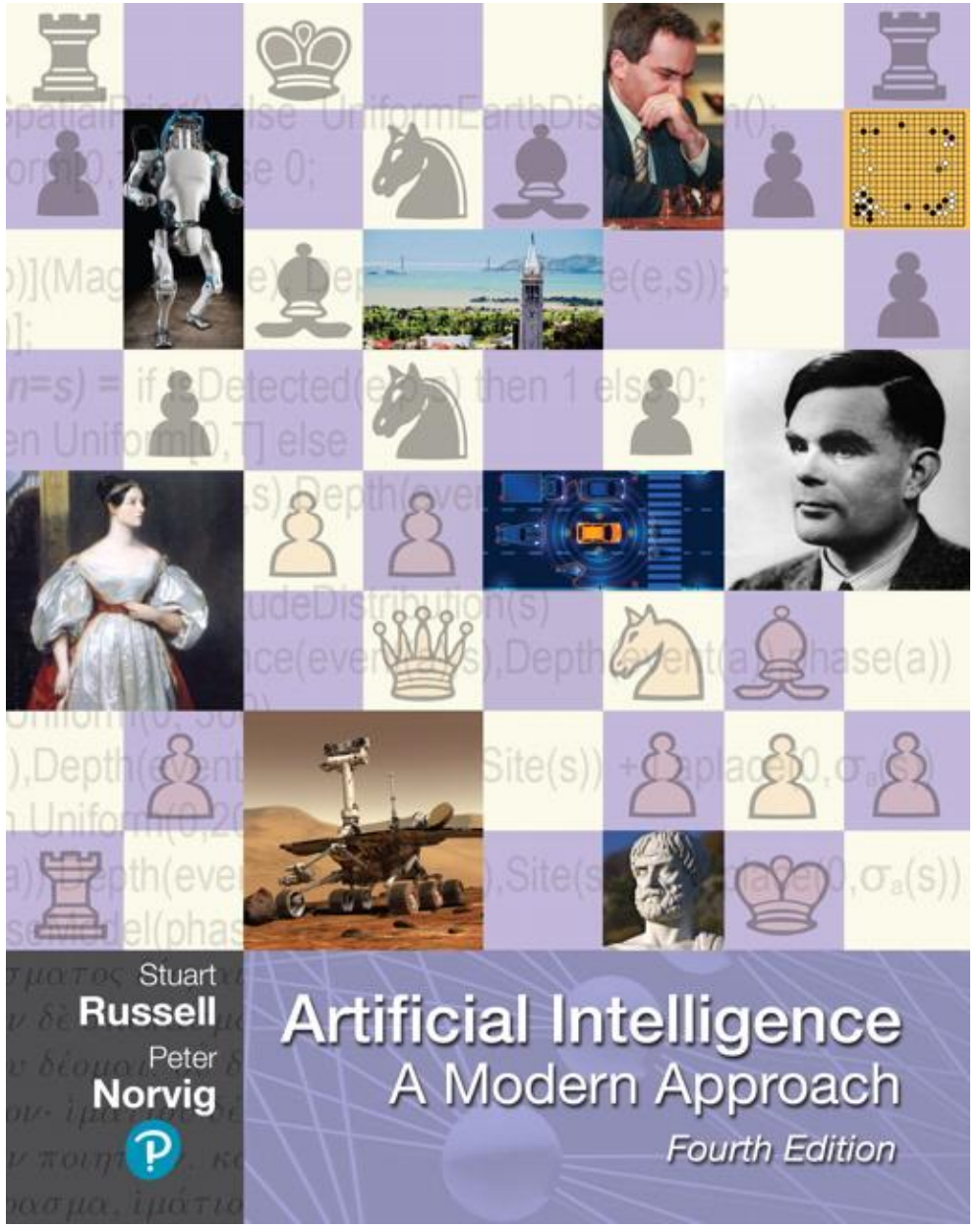
Outline for today's lecture

Intelligent Agents (AIMA 2.1-2.4)

Task Environments

Formulating Search Problems

Uninformed Search (AIMA 3.1-3.4)



Four views of Artificial Intelligence

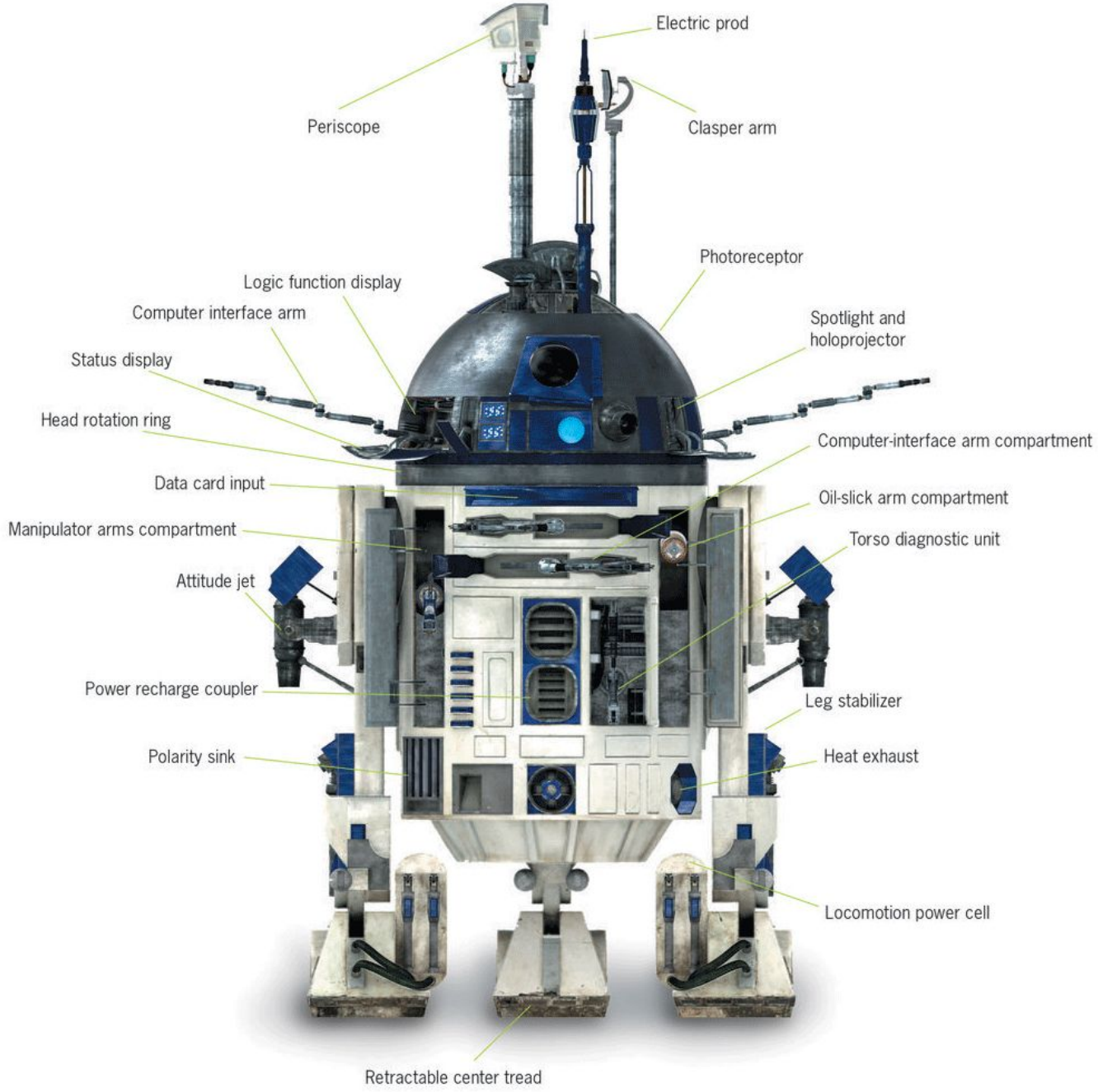
Thinking humanly	**Thinking rationally**
Acting humanly	**Acting rationally**

This course is about effective programming techniques for designing rational agents

Agents

An **agent** is anything that **perceives** its environment through **sensors** and can **act** on its environment through **actuators**

A **percept** is the agent's perceptual inputs at any given instance.

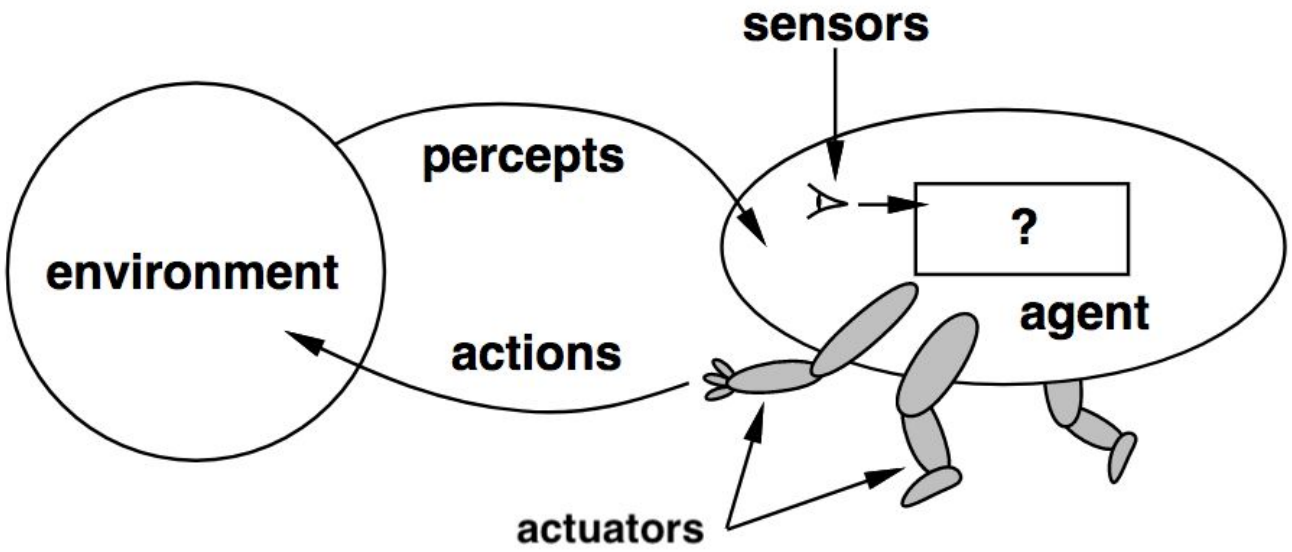


What about your robot?

What actuators does it have?

What sensors does it have?

Agents and environments



An agent is specified by an *agent function* $f:P \rightarrow a$ that maps a sequence of percept vectors P to an action a from a set A:

$P=[p_0, p_1, \ldots, p_t]$

$A=\{a_0, a_1, \ldots, a_k\}$

abstract mathematical description

Agent function & program

The ***agent program*** runs on the physical ***architecture*** to *produce* f

- *agent* = *architecture* + *program*

"Easy" solution: a giant table that maps every possible sequence P to an action a

- One problem: exponential in length of P

Agents

An ***agent*** is anything that can be viewed as

- ***perceiving*** its ***environment*** through ***sensors*** and
- ***acting*** upon that environment through ***actuators***

Human agent:

- Sensors: eyes, ears, …
- Actuators: hands, legs, mouth, …

Robotic agent:

- Sensors: cameras and infrared range finders
- Actuators: various motors

Agents include humans, robots, softbots, *thermostats*, …

Acting rationally

Rational behavior: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Rational agent: An agent is an entity that perceives and acts rationally

This course is about effective programming techniques for designing rational agents

Rational Behavior

Rational Agent

Let's try to define "rational agent".

A **rational agent** is an agent that perceives its environment and and behaves rationally

Rational behavior: doing the right thing

Obviously doing the right thing is better than doing the wrong thing, but *what does it mean to do the right thing?*

In Philosophy

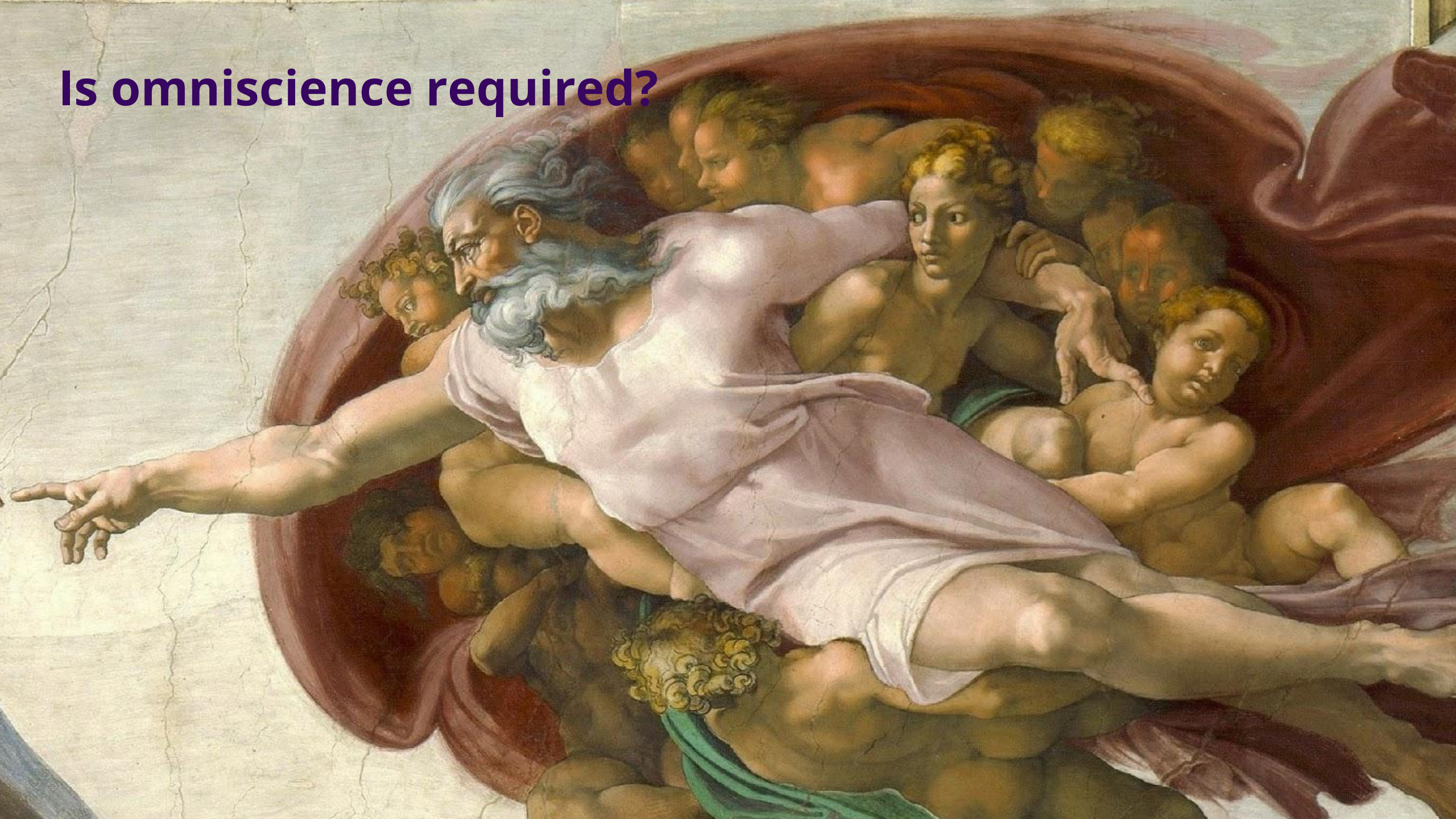
Moral philosophy has developed different notions of “the right thing”.

AI is usually concerned with **Consequentialism.**

We evaluate an agent’s behavior by its consequences.

The Good Place

Is omniscience required?



In Economics

A BEHAVIORAL MODEL OF RATIONAL CHOICE

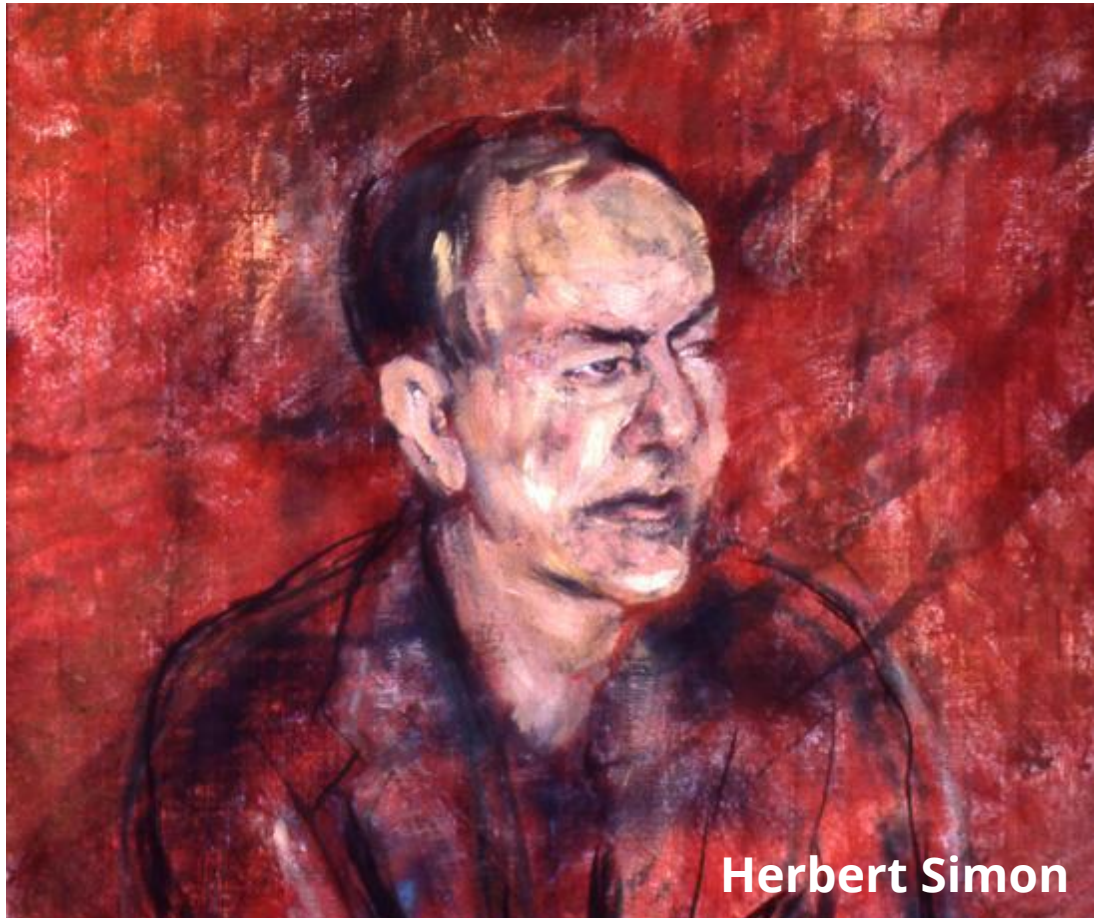
Herbert A. Simon*

Summary: A model is proposed for the description of rational choice by organisms of limited computational ability.

Rational choice theory is a framework for understanding social and economic behavior.

The basic premise is that aggregate social behavior results from the behavior of individual actors, each of whom is making their individual decisions.

It assumes that individuals have preferences and choose the alternative that they prefer.



Herbert Simon

Performance measure

How do we know if an agent is acting rationally?

- Informally, we expect that it will do the right thing in all circumstances.

How do we know if it's doing the right thing?

We define a **performance measure:**

- An objective criterion for success of an agent's behavior
- given the evidence provided by the percept sequence.

Performance measure examples

Performance measures for a vacuum-cleaner agent might include things like:

- +1 point for each clean square at time T
- +1 for cleaning a square, -1 for each move
- -1000 for more than *k* dirty squares

Rule of thumb for performance measures

It is better to design performance measures according to **what you want to be achieved** in the environment, **rather than how you think the agent should behave.**

For example, what might happen if we said

- +1 point for each time the robot cleans a square

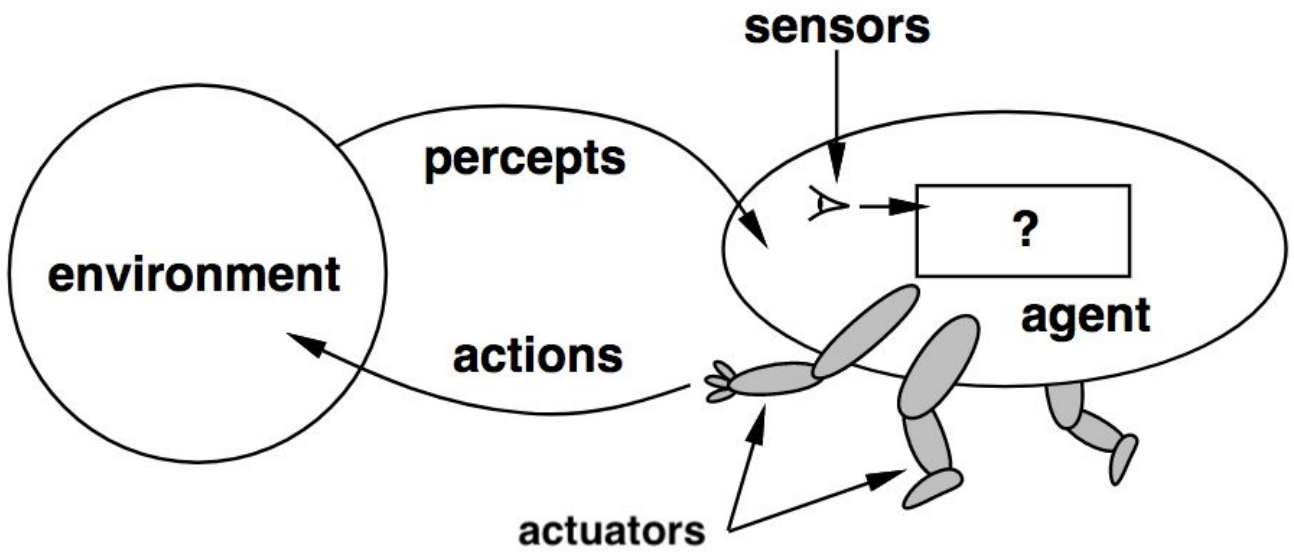
instead of

- +1 point for each clean square at time T

Rational agents

Rational Agent:

- For each possible percept sequence *P*
- a rational agent selects an action *a*
- to ***maximize*** its ***performance measure***



Rationality is *not* omniscience

Ideal agent: maximizes ***actual*** performance, but needs to be ***omniscient***.

- Usually impossible…..
  - But consider tic-tac-toe agent…
- Rationality ≠ Guaranteed Success

Caveat: computational limitations make ***complete rationality*** unachievable

- design best ***program*** for given machine resources

Expected value

Rational Agent (initial definition):

- For each possible percept sequence P,
- a rational agent selects an action a
- to maximize its performance measure

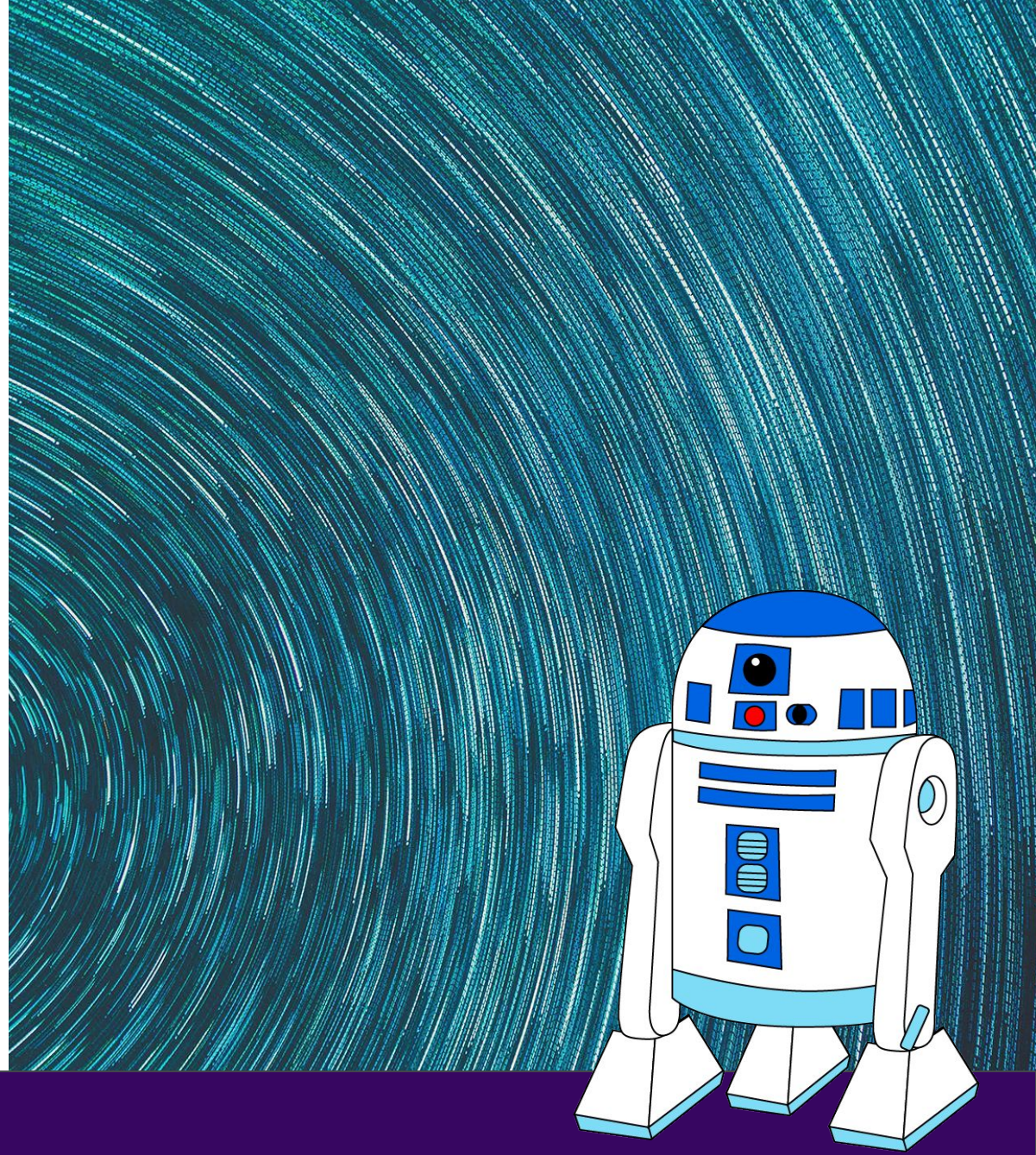
Rational Agent (revised definition):

- For each possible percept sequence P,
- a rational agent selects an action a
- that maximizes the **expected value** of its performance measure

It doesn't have to know what the actual outcome will be.

CIS 421/521:
ARTIFICIAL INTELLIGENCE

Task Environments



Penn Engineering
University of Pennsylvania

Task environments

To design a rational agent we need to specify a ***task environment***

- a problem specification for which the agent is a solution

PEAS: *to* specify a task environment

- ***P***erformance measure
- ***E***nvironment
- ***A***ctuators
- ***S***ensors

PEAS: Specifying an automated taxi driver

Performance measure:

- ?

Environment:

- ?

Actuators:

- ?

Sensors:

- ?

PEAS: Specifying an automated taxi driver

Performance measure:

- safe, fast, legal, comfortable, maximize profits

Environment:

- roads, other traffic, pedestrians, customers

Actuators:

- steering, accelerator, brake, signal, horn

Sensors:

- cameras, LiDAR, speedometer, GPS

https://www.today.com/video/amazon-adebuts-new-package-delivery-drone-61414981780

PEAS: Amazon Prime Air

Performance measure:

- ?

Environment:

- ?

Actuators:

- ?

Sensors:

- ?

PEAS: Specifying an Amazon delivery drone

*P*erformance measure:

maximize profits - minimize time - obey laws governing airspace restrictions - deliver package to right location - keep package in good condition - avoid accidents - reduce noise - preserve battery life

*E*nvironment:

- airspace - obstacles when airborne (other drones, birds, buildings, trees, utility poles) - obstacles when landing (pets, patio furniture, lawnmowers, people, cars) - weather - distances/route information between warehouse and destinations - position of houses, and spaces that are safe for drop-off- package weight

PEAS: Specifying an Amazon delivery drone

*A*ctuators:

- Propellers and flight control system- Payload actuators: E.g. Arm/basket/claw for picking up, dropping off packages- Lights or signals - Mechanism to announce/verify delivery- Device for delivering packages to customers

*S*ensors:

- GPS - radar/Lidar- altitude sensor- weather sensors (barometer, etc). - gyroscope- accelerometer- camera- rotor sensors- weight sensor to recognize package

The rational agent designer's goal

Goal of AI practitioner who designs rational agents:

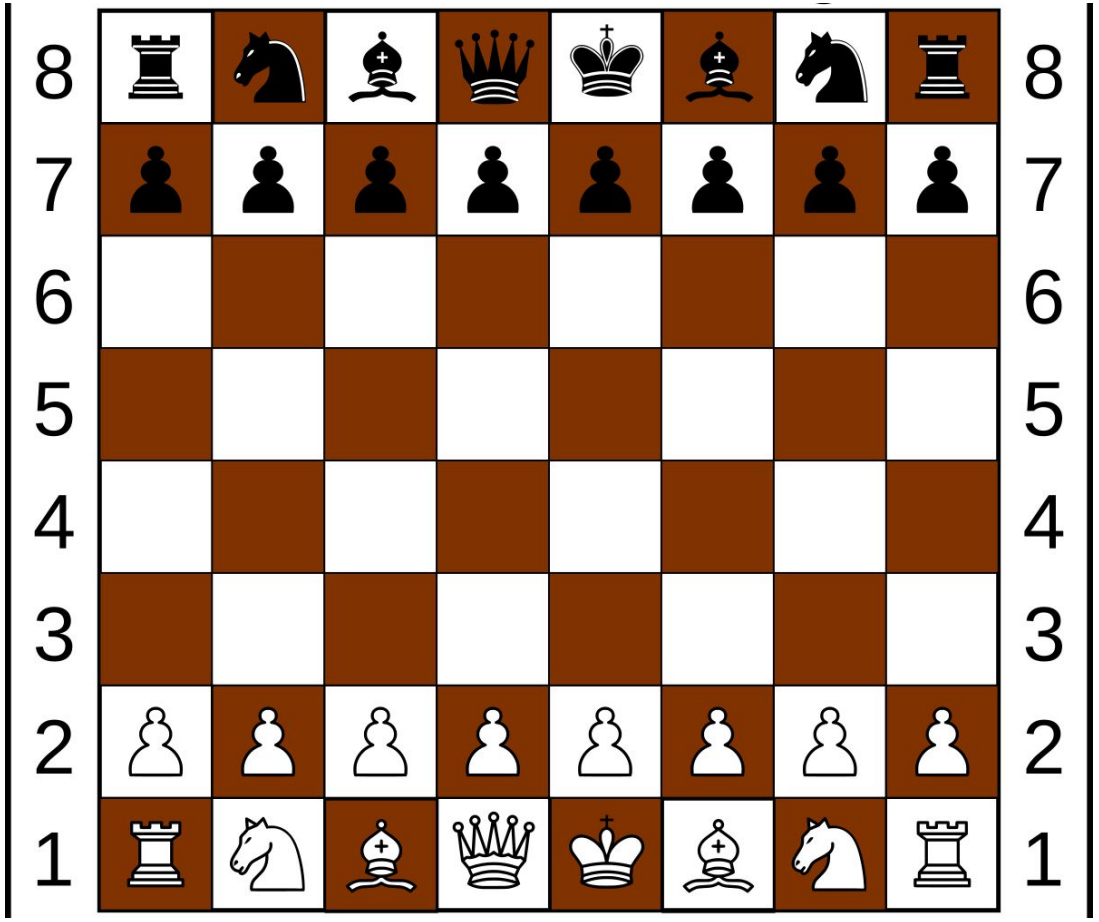
given a *PEAS* task environment,

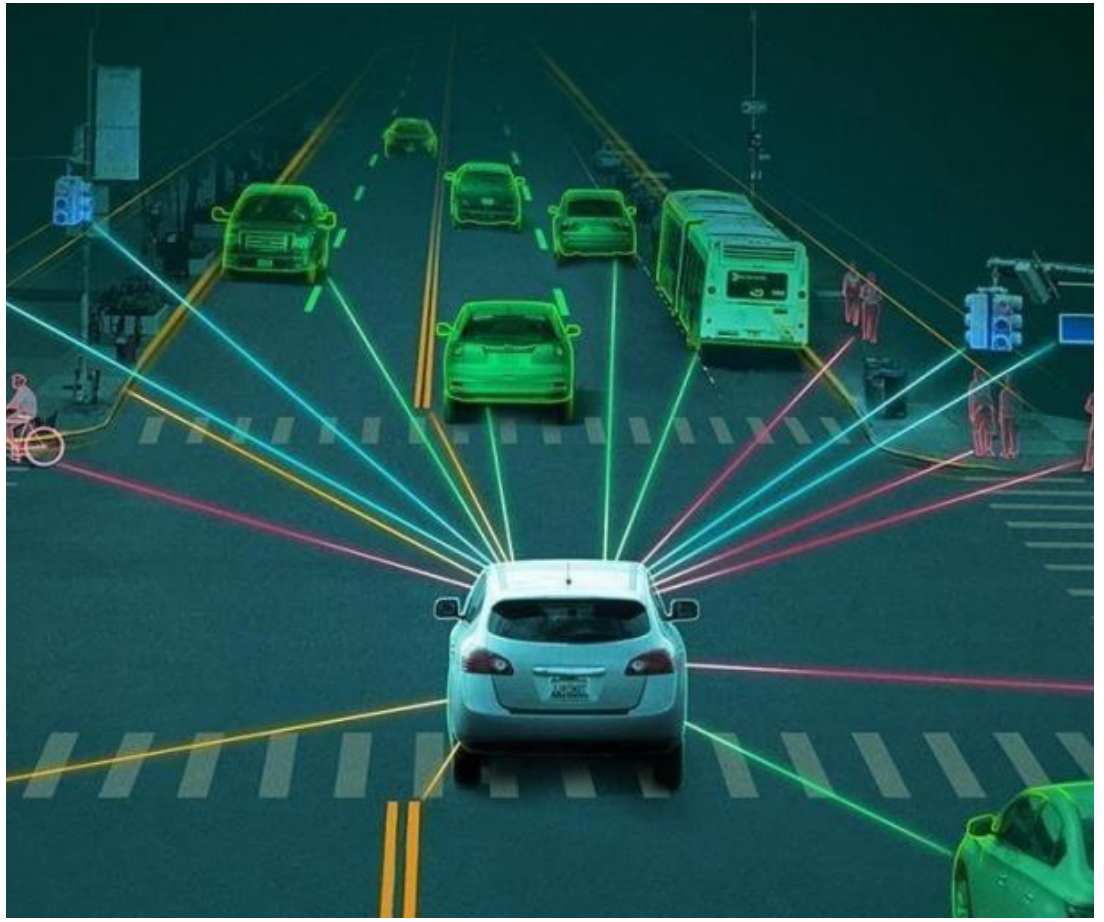
1. **Construct *agent function* f that maximizes the expected value of the performance measure,**
2. **Design an *agent program* that implements f on a particular architecture**

abstract mathematical description

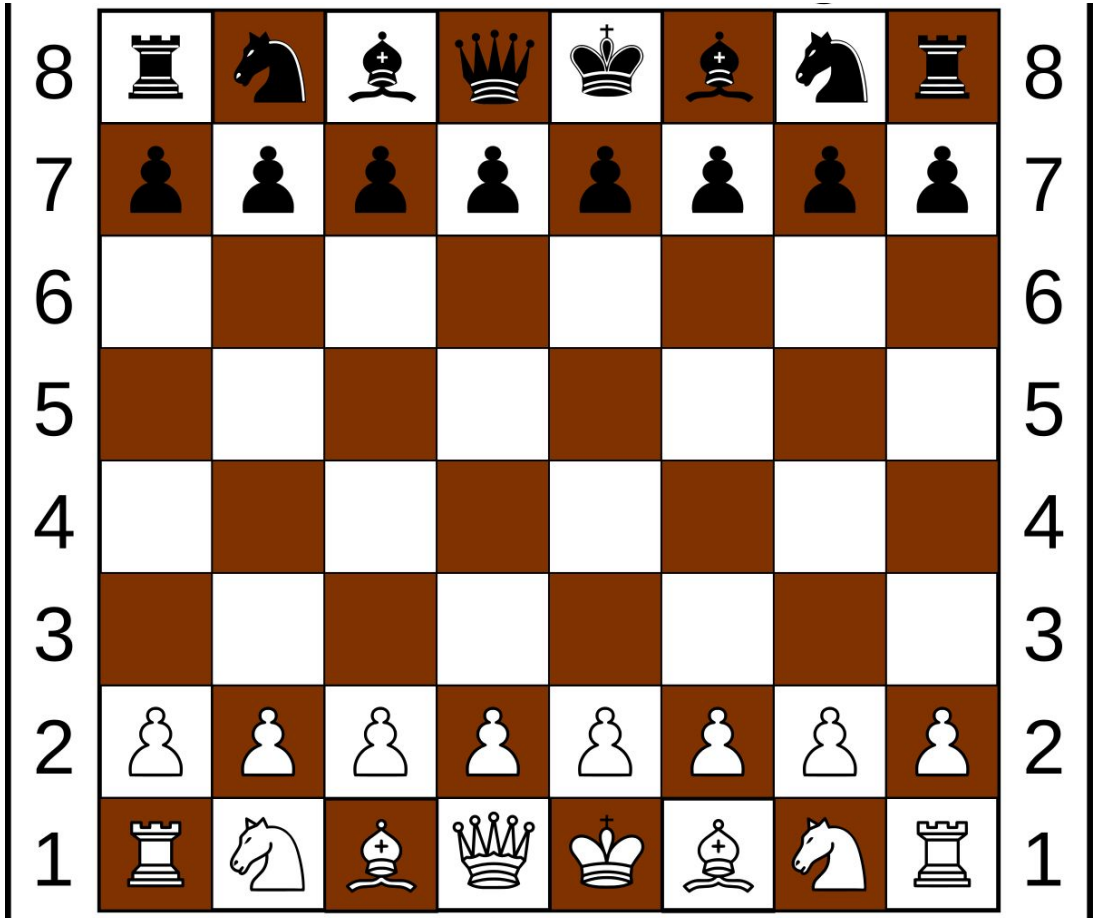
concrete implementation

Fully Observable v. Partially Observable

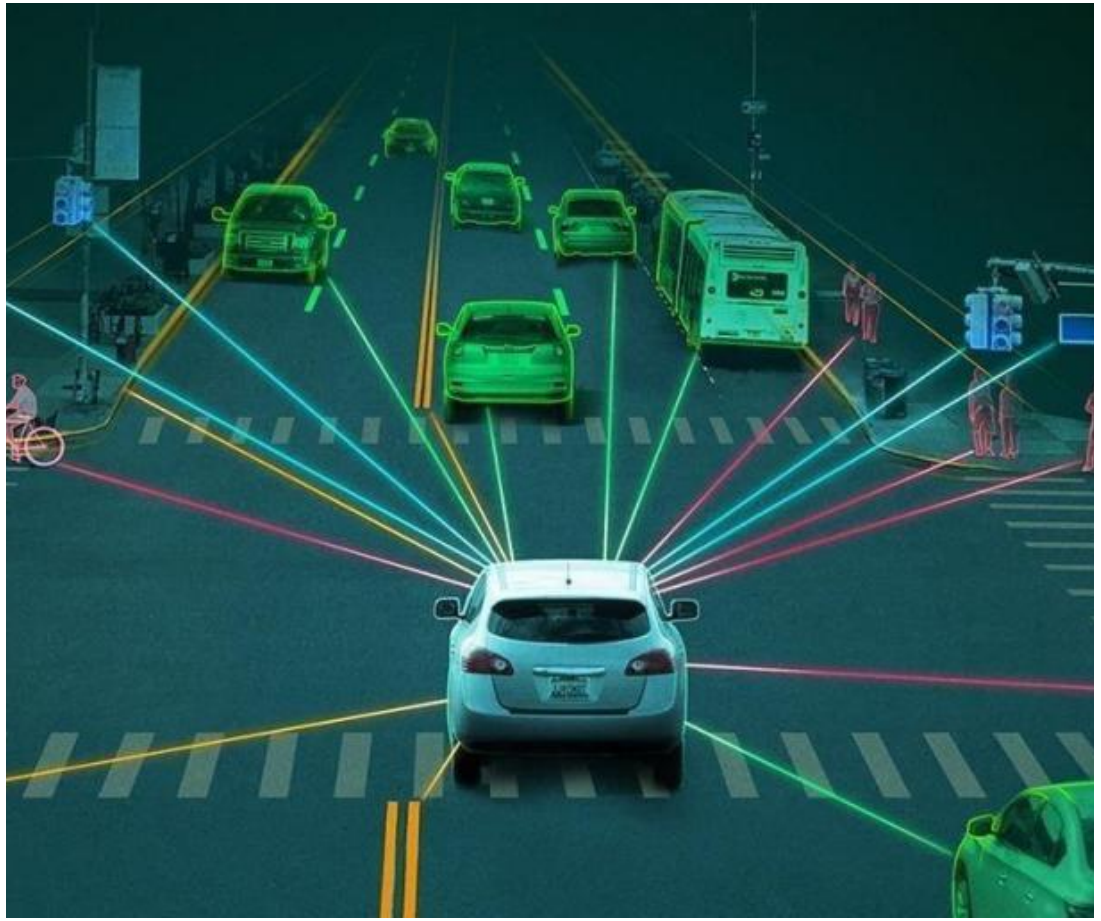




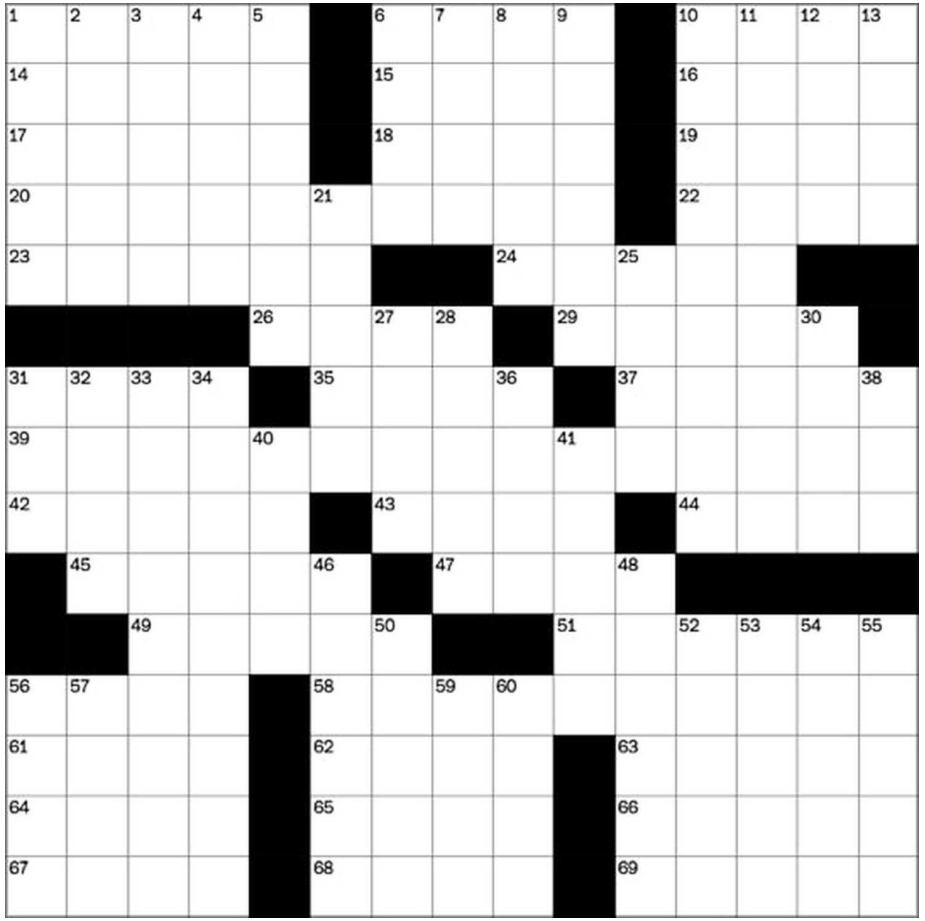
Deterministic v. Nondeterministic v. Stochastic

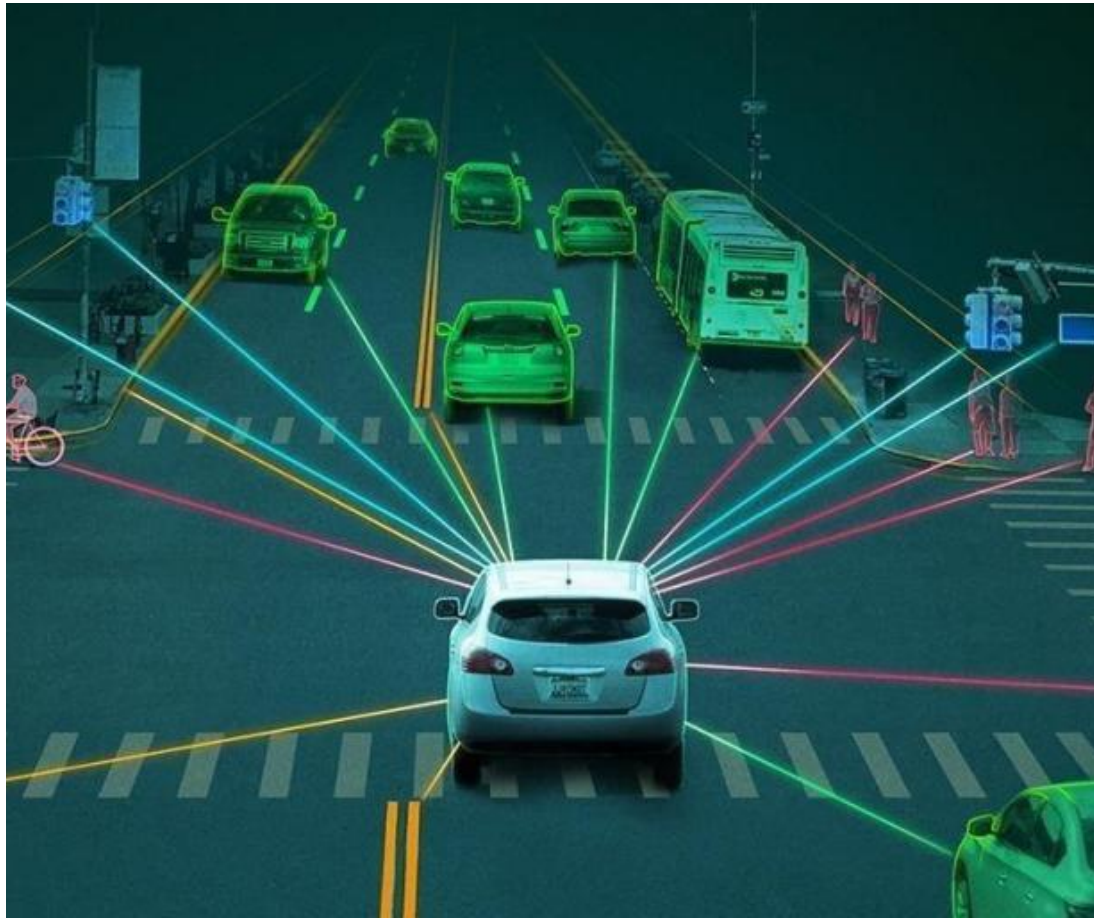


Episodic v. Sequential

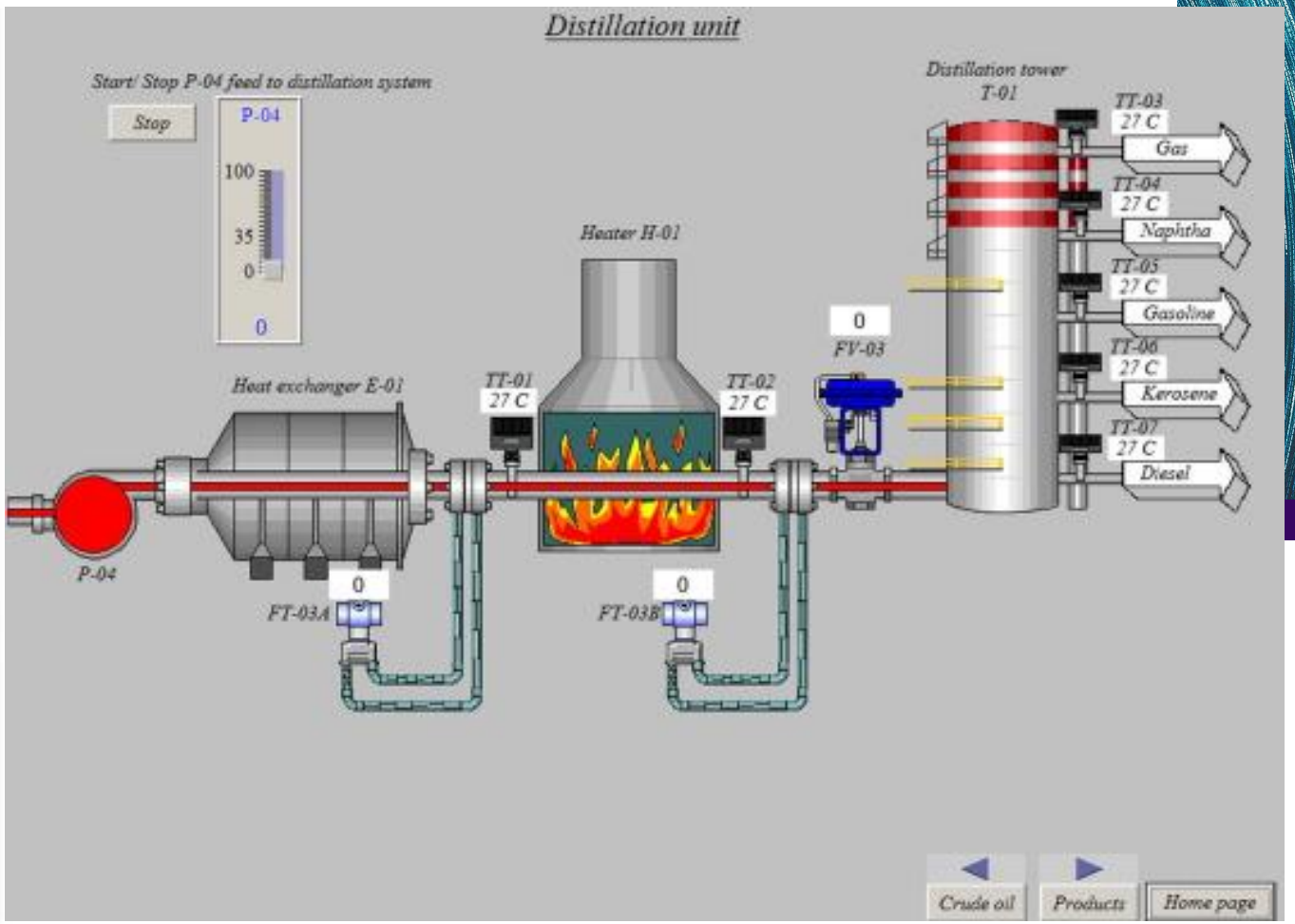


Static v. Dynamic

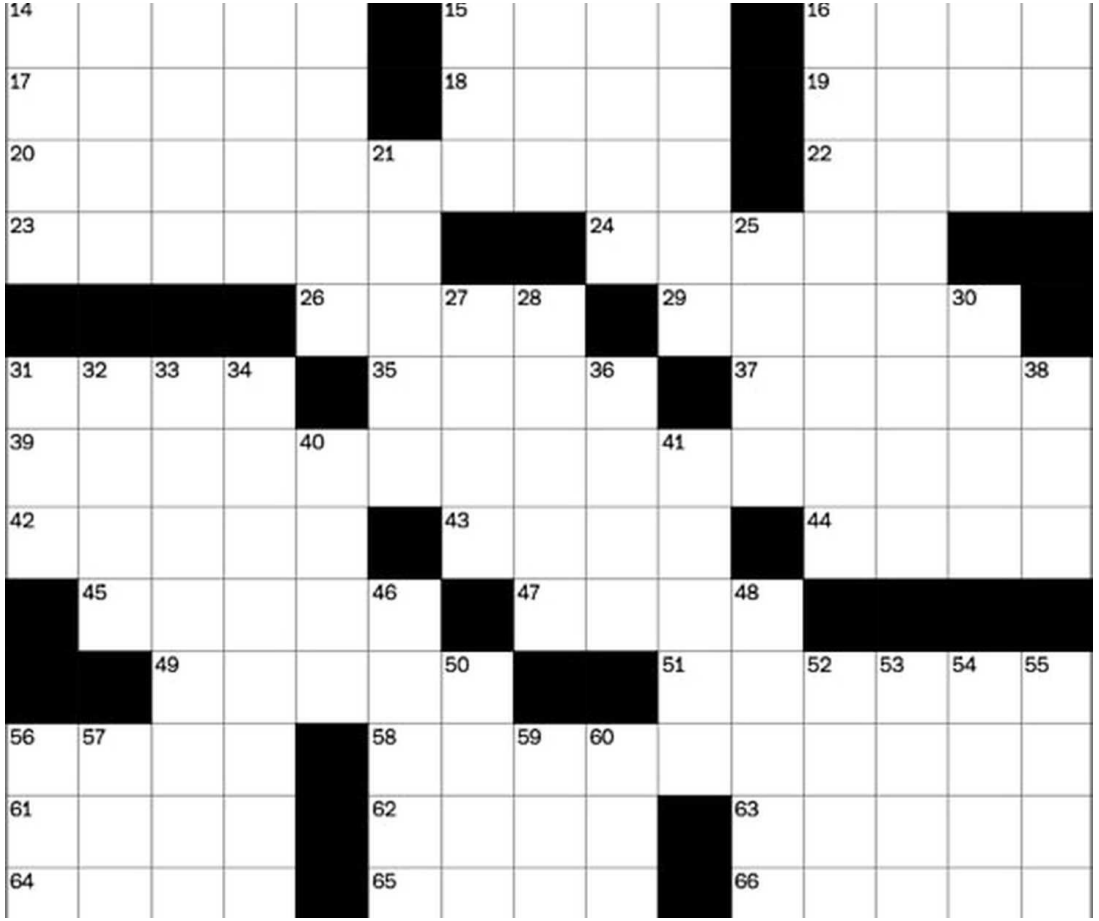


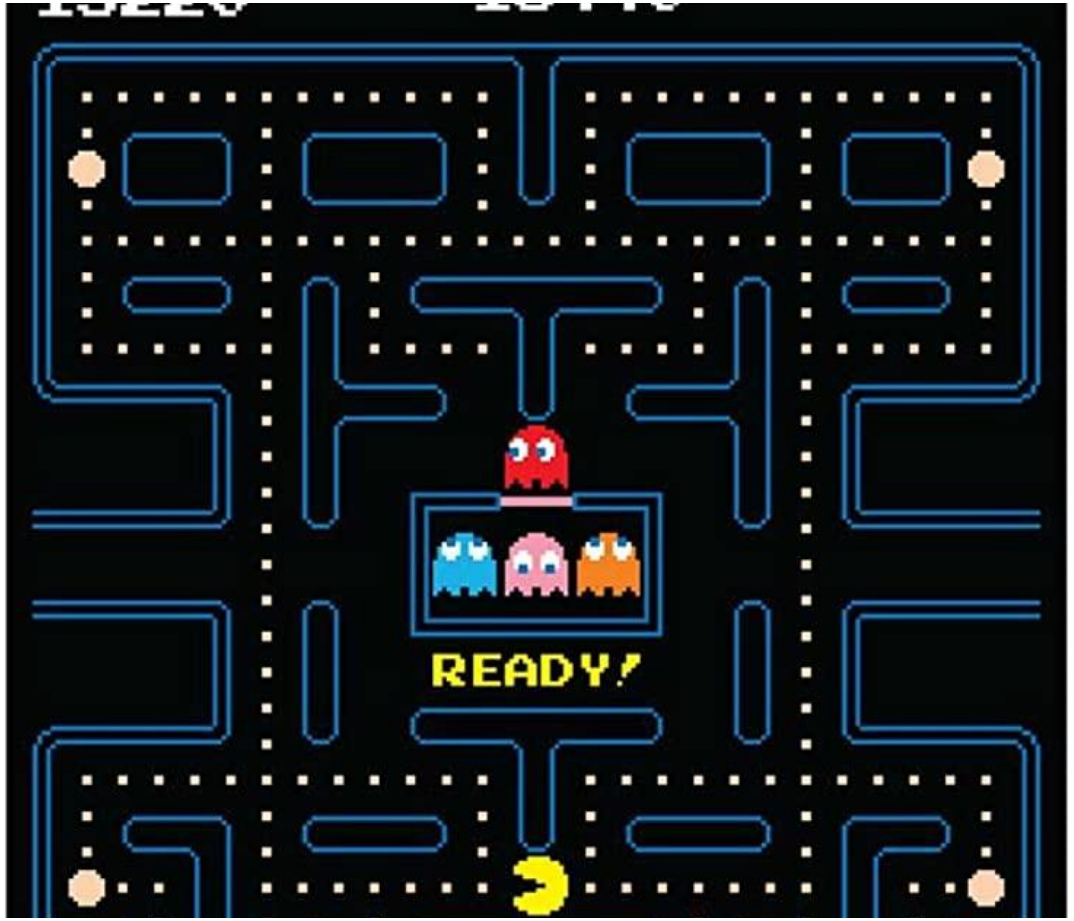


Discrete v. Continuous



Single Agent v. Multi Agent





When should something be considered an agent?

When should something be considered another agent?

If we're talking about a self driving taxi, when should we consider something part of the environment versus another agent?

For instance, a telephone pole is part of the environment, but a car might be another agent.

When something behavior can best be described as having its own performance measure, then we should consider it to be an agent.

Examples

TASK ENVIRONMENT	OBSERVABLE	AGENTS	DETERMINISTIC	EPISODIC	STATIC	DISCRETE
Crossword puzzle	Fully	Single	Deterministic	Sequential	Static	Discrete
Chess with a clock	Fully	Multi	Deterministic	Sequential	Semi	Discrete
Poker	Partially	Multi	Stochastic	Sequential	Static	Discrete
Backgammon	Fully	Multi	Stochastic	Sequential	Static	Discrete
Taxi driving	Partially	Multi	Stochastic	Sequential	Dynamic	Continuous
Medical diagnosis	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Image analysis	Fully	Single	Deterministic	Episodic	Semi	Continuous
Part-picking robot	Partially	Single	Stochastic	Episodic	Dynamic	Continuous
Refinery controller	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Interactive English tutor	Partially	Multi	Stochastic	Sequential	Dynamic	Discrete

The Hardest Environment

The hardest case is:

- **CONTINUOUS**
- **PARTIALLY OBSERVABLE**
- **STOCHASTIC**
- **MULTIAGENT**
- **UNKNOWN OUTCOMES**

Environment Restrictions for Now

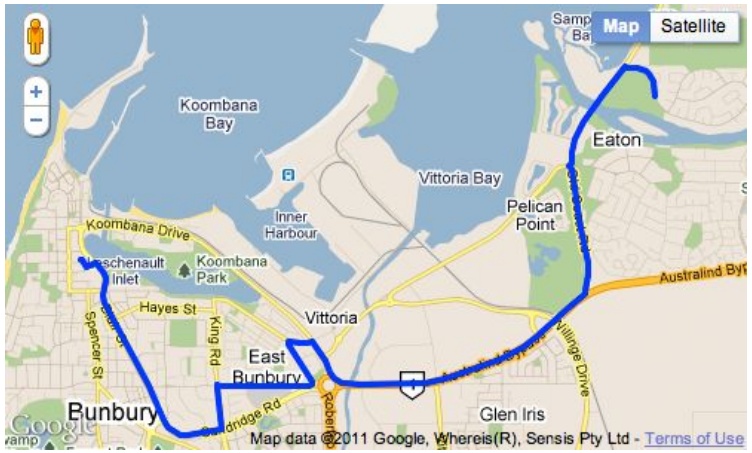
We will assume environment is:

- **STATIC**
- **FULLY OBSERVABLE**
- **DETERMINISTIC**
- **DISCRETE**

Reflex agents v. Problem solving agents

A simple reflex agent is one that selects an action based on the current percept, and ignores the rest of the percept history.

A problem-solving agent must plan ahead. It will consider a sequence of actions that form a path to a goal state. The computational process that it undertakes is called search.



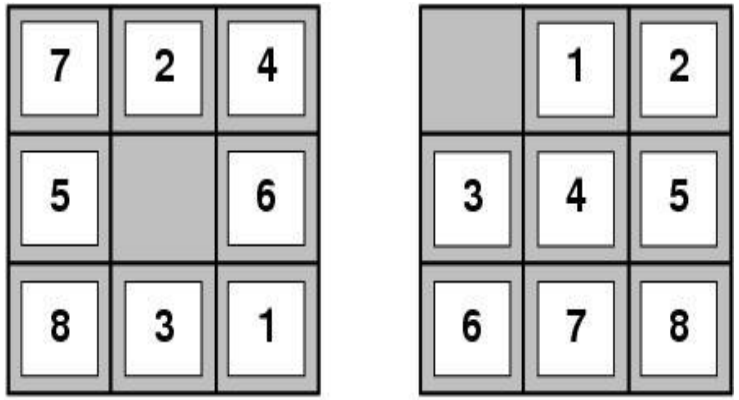
Problem Solving Agents & Problem Formulation

AIMA 3.1-3.2

Example search problem: 8-puzzle

Formulate ***goal***

- Pieces to end up in order as shown…



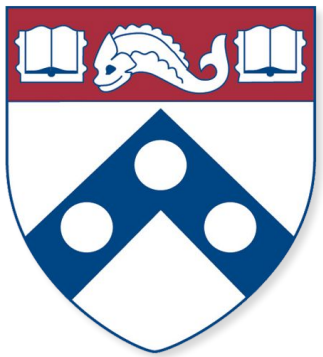
Start State Goal State

Formulate ***search problem***

- ***States:*** configurations of the puzzle (9! configurations)
- ***Actions***: Move one of the movable pieces (≤4 possible)
- ***Performance measure***: minimize total moves

Find ***solution***

- Sequence of pieces moved: 3,1,6,3,1,…



Penn
Engineering

Copyright 2020 University of Pennsylvania
No reproduction or distribution without permission.