

At the end of the class you should be able to:

- characterize simplifying assumptions made in building AI systems
- determine what simplifying assumptions particular AI systems are making
- suggest what assumptions to lift to build a more intelligent system than an existing one

- Research proceeds by making simplifying assumptions, and gradually reducing them.
- Each simplifying assumption gives a dimension of complexity
  - ▶ multiple values in a dimension: from simple to complex
  - ▶ simplifying assumptions can be relaxed in various combinations

# Dimensions of complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

- Model at one level of abstraction: **flat**
- Model with interacting modules that can be understood separately: **modular**
- Model with modules that are (recursively) decomposed into modules: **hierarchical**
- **Example:** Planning a trip from here to a see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical
- A flat description is either continuous or discrete. Hierarchical reasoning is often a hybrid of continuous and discrete.

...how far the agent looks into the future when deciding what to do.

- **Static:** world does not change
- **Finite stage:** agent reasons about a fixed finite number of time steps
- **Indefinite stage:** agent reasons about a finite, but not predetermined, number of time steps
- **Infinite stage:** the agent plans for going on forever (process oriented)

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.

A agent can reason in terms of:

- **Explicit states** — a state is one way the world could be
- **Features** or **propositions**.
  - ▶ States can be described using features.
  - ▶ 30 binary features can represent  $2^{30} = 1,073,741,824$  states.
- **Individuals** and **relations**
  - ▶ There is a feature for each relationship on each tuple of individuals.
  - ▶ Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

- **Perfect rationality:** the agent can determine the best course of action, without taking into account its limited computational resources.
- **Bounded rationality:** the agent must make good decisions based on its perceptual, computational and memory limitations.

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

... always some mix of prior (innate, programmed) knowledge and learning (nature vs nurture).

- Learning is impossible without prior knowledge (bias).



There are two dimensions for uncertainty. In each dimension an agent can have

- **No uncertainty:** the agent knows what is true
- **Disjunctive uncertainty:** there is a set of states that are possible
- **Probabilistic uncertainty:** a probability distribution over the worlds.

# Why probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - ▶ definitive predictions: you will be run over tomorrow
  - ▶ disjunctions: be careful or you will be run over
  - ▶ point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Whether an agent can determine the state from its stimuli:

- **Fully-observable**: the agent can observe the state of the world.
- **Partially-observable**: there can be a number states that are possible given the agent's stimuli.

If an agent knew the initial state and its action, could it predict the resulting state?

The dynamics can be:

- **Deterministic**: the resulting state is determined from the action and the state
- **Stochastic**: there is uncertainty about the resulting state.

What does the agent try to achieve?

- **achievement goal** is a goal to achieve. This can be a complex logical formula.
- **complex preferences** may involve tradeoffs between various desiderata, perhaps at different times.
  - ▶ **ordinal** only the order matters
  - ▶ **cardinal** absolute values also matter

**Examples:** coffee delivery robot, medical doctor

Are there multiple reasoning agents that need to be taken into account?

- **Single agent** reasoning: any other agents are part of the environment.
- **Multiple agent** reasoning: an agent reasons strategically about the reasoning of other agents.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

When does the agent reason to determine what to do?

- **reason offline**: before acting
- **reason online**: while interacting with environment

# Dimensions of complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online



# State-space search

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Deterministic planning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Decision networks

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Markov decision processes (MDPs)

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Decision-theoretic planning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Reinforcement learning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

# Classical game theory

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Dimension	Values
Modularity	flat, modular, <b>hierarchical</b>
Planning horizon	non-planning, finite stage, <b>indefinite stage, infinite stage</b>
Representation	states, features, <b>relations</b>
Computational limits	perfect rationality, <b>bounded rationality</b>
Learning	knowledge is given, <b>knowledge is learned</b>
Sensing uncertainty	fully observable, <b>partially observable</b>
Effect uncertainty	deterministic, <b>stochastic</b>
Preference	goals, <b>complex preferences</b>
Number of agents	single agent, <b>multiple agents</b>
Interaction	offline, <b>online</b>



# The dimensions interact in complex ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:
  - ▶ Hierarchical reasoning
  - ▶ Individuals and relations
  - ▶ Bounded rationality