Dimensions

- 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

- Flat or modular or hierarchical
- Explicit states or features or individuals and relations
- Static or finite stage or indefinite stage or infinite stage
- Fully observable or partially observable
- Deterministic or stochastic dynamics
- Goals or complex preferences
- Single-agent or multiple agents
- Knowledge is given or knowledge is learned from experience
- Perfect rationality or bounded rationality

Modularity

- 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 resort in Cancun, Mexico
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical

Succinctness and Expressiveness

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

Succinctness and Expressiveness

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.

Succinctness and Expressiveness

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.

Representation of states

- A thermostat for a heater has two belief states: off and heating.
- The environment has three states: cold, comfortable, and hot.
- There are thus six states corresponding to the different combinations of belief and environment states.
- These states may not fully describe the world, but they are adequate to describe what a thermostat should do. The thermostat should move to, or stay in, heating if the environment is cold and move to, or stay in, off if the environment is hot. If the environment is comfortable, the thermostat should stay in its current state. The agent heats in the heating state and does not heat in the off state.

States as features

A proposition is a Boolean feature, which means that its value is either true or false.

Thirty propositions can encode $2^{30} = 1,073,741,824$ states.

- An agent that has to look after a house may have to reason about whether light bulbs are broken.
- It may have features for the position of each switch, the status of each switch (whether it is working okay, whether it is shorted, or whether it is broken), and whether each light works.
- The feature pos-s2 may be a feature that has value up when switch s2 is up and has value down when the switch is down.
- The state of the house's lighting may be described in terms of values for each of these features.

States as relations

It is often more convenient to have relational descriptions in terms of individuals and relations among them.
One binary relation and 100 individuals can represent $100^2 = 10,000$ propositions and $2^{10,000}$ states.

- The agent that looks after a house in the previous could have the lights and switches as individuals, and relations position and connected-to.
- Instead of the feature position-s1=up, it could use the relation position(s1,up).
- This relation enables the agent to reason about all switches or for an agent to have knowledge about switches that can be used when the agent encounters a switch.

Planning horizon

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

- Static: world does not change
- Finite stage: agent reasons on a fixed number of time steps
  - For example, a doctor may have to treat a patient but may have time for some testing and so there may be two stages: a testing stage and a treatment stage to plan for.
  - In the degenerate case where an agent only looks one time step ahead, it is said to be *greedy* or myopic
- Indefinite stage: agent reasons about a finite, but not predetermined, number of time steps
  - For example, an agent that must get to some location may not know a priori how many steps it will take to get there.
- Infinite stage: the agent goes on forever (process oriented)
  - For example, the stabilization module of a legged robot should go on forever; it cannot stop when it has achieved stability, because the robot has to keep from falling over.

Uncertainty

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

- **No uncertainty:** the agent knows which world is true
- **Disjunctive uncertainty:** there is a set of worlds 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
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Sensing Uncertainty

Whether an agent can determine the state from its observations:

- 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 observations.

Uncertain dynamics

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.

Goals or complex preferences

- achievement goal is a goal to achieve. This can be a complex logical formula.
  - For example, the goals for a robot may be to get two cups of coffee and a banana, and not to make a mess or hurt anyone.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters
    - For example, Sam prefers cappuccino over black coffee and prefers black coffee over tea.
  - cardinal absolute values also matter
    - For example, a trade-off between the wait time and the type of beverage, and a mess-taste trade-off, where Sam is prepared to put up with more mess in the preparation of the coffee if the taste of the coffee is exceptionally good.

Single agent or multiple agents

- Single agent reasoning is where an agent assumes that any other agents are part of the environment.
- Multiple agent reasoning is when 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

Learning from experience

Whether the model is fully specified a priori:

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

Perfect rationality or bounded rationality

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

Dimensions of Complexity

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single-agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

State-space Search

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Classical Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Decision Networks

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Markov Decision Processes (MDPs)

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Decision-theoretic Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Reinforcement Learning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Classical Game Theory

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

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