# Contents

# 1 Designing intelligent agents

- an *agent* operates in a task environment:
  - *task*: the goal(s) the agent is trying to achieve
  - *environment*: that part of the real world or a computational system inhabited by the agent
- agent obtains information about the environment in the form of percepts
- agent changes the environment by performing actions to achieve its goals

# 2 From agent functions to agent programs

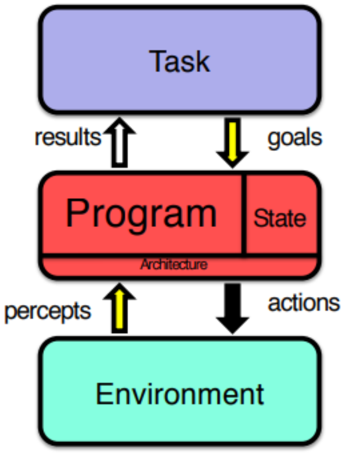
- the **behaviour** of an agent is described by an *agent function* (action selection function) which maps a goal and sequence of percepts to an action (and possibly results)
- agent programming is conventionally conceived as the problem of synthesising an agent function
- Difficult to design, because of unknown environments.

# 3 Agent architectures

- one way of making the agent programming problem more tractable is make use of the notion of an agent architecture
- the notion of an agent architecture is ubiquitous in the agent literature but is not well analysed
- often discussed in the context of an agent programming language or platform
- architecture is a blueprint for software agents and intelligent control systems, depicting the arrangement of components

## 3.1 Russell and Norvig view of an agent

- **program**: implements the agent function mapping from goals and percepts to actions (and results)
- **state**: includes all the internal representations on which the agent program operates
- **architecture**: computing device with sensors and actuators that runs the agent program



## 3.2 Architecture as a virutal machine

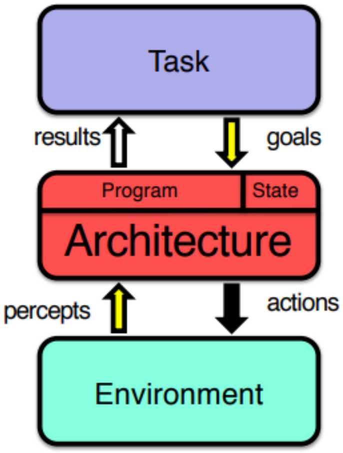
- the architecture defines a (real or virtual) machine which runs the agent program
- defines the atomic operations of the agent program and implicitly determines the components of the agent
- determines which operations happen automatically, without the agent program having to do anything
- e.g., the interaction between memory, learning and reasoning
- an architecture constrains kinds of agent programs we can write (easily)

### 3.3 Architectural view of an agent

- **program**: a function mapping from goals and percepts to actions (and results) expressed in terms of virtual machine operations

- **state**: the virtual machine representations on which the agent program operates

- **architecture**: a virtual machine that runs the agent program and updates the agent state



### 3.4 Hierarchies of virtual machines

In many agents we have a whole hierarchy of virtual machines, used without qualification, agent architecture means the most abstract architecture or the highest level virtual machine

## 4 Cognitive architecture

- agent architecture is also related to the notion of a cognitive architecture as used in artificial intelligence and cognitive science
- a *cognitive architecture* is an integrated system capable of supporting intelligence
- often used to denote models of **human reasoning**, e.g., ACT-R, SOAR
- in other cases no claims about psychological plausibility are made
- in this latter sense, cognitive architecture is more or less synonymous with agent architecture as used here

### 4.1 Properties of the architecture

- an agent architecture can be seen as defining a class of agent programs
- just as individual agent programs have properties that make them more or less successful in a given task environment
- architectures (classes of programs) have higher-level properties that determine their suitability for a task environment
- choosing an appropriate architecture can make it much easier to develop an agent program for a particular task environment

## Reference section

**agent function**

The agent function is a mathematical function that maps a sequence of perceptions into action

**agent**

In artificial intelligence, an intelligent agent (IA) is an autonomous **entity**, which observes through sensors and acts upon an environment using actuators and directs its activity towards achieving goal