INTRODUCTION TO AGENTS: WHAT THEY ARE/HOW THEY WORK

Some text and images in these slides were drawn from
Russel & Norvig's published material

WHAT IS ARTIFICIAL INTELLIGENCE?

2

- Definitions of AI vary
- Artificial Intelligence is the study of systems that

think like humans	think rationally
act like humans	act rationally

SYSTEMS ACTING LIKE HUMANS

3

- Turing test: test for intelligent behavior
  - Interrogator writes questions and receives answers
  - System providing the answers passes the test if interrogator cannot tell whether the answers come from a person or not
- Necessary components of such a system form major AI sub-disciplines:
  - Natural language, knowledge representation, automated reasoning, machine learning

SYSTEMS THINKING LIKE HUMANS

4

- Formulate a theory of mind/brain
- Express the theory in a computer program
- Two Approaches
  - Cognitive Science and Psychology (testing/ predicting responses of human subjects)
  - Cognitive Neuroscience (observing neurological data)

5 SYSTEMS THINKING RATIONALLY

- "Rational" -> **ideal** intelligence
  (contrast with **human** intelligence)
- Rational thinking governed by precise "laws of thought"
  - syllogisms
  - notation and logic
- Systems (in theory) can solve problems using such laws

SYSTEMS ACTING RATIONALLY

- Building systems that carry out actions to achieve the **best outcome**
- Rational **behavior**
- May or may not involve rational thinking
  - i.e., consider reflex actions
- This is the definition we will adopt

7 INTELLIGENT AGENTS

- Agent: anything that perceives and acts on its environment
- AI: study of rational agents
- A rational agent carries out an action with the **best outcome** after considering past and current percepts

FOUNDATIONS OF AI

- Philosophy: logic, mind, knowledge
- Mathematics: proof, computability, probability
- Economics: maximizing payoffs
- Neuroscience: brain and neurons
- Psychology: thought, perception, action
- Control Theory: stable feedback systems
- Linguistics: knowledge representation, syntax

BRIEF HISTORY OF AI

- 1943: McCulloch & Pitts: Boolean circuit model of brain
- 1950: Turing's "Computing Machinery and Intelligence"
- 1952—69: Look, Ma, no hands!
- 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
- 1956: Dartmouth meeting: "Artificial Intelligence" adopted

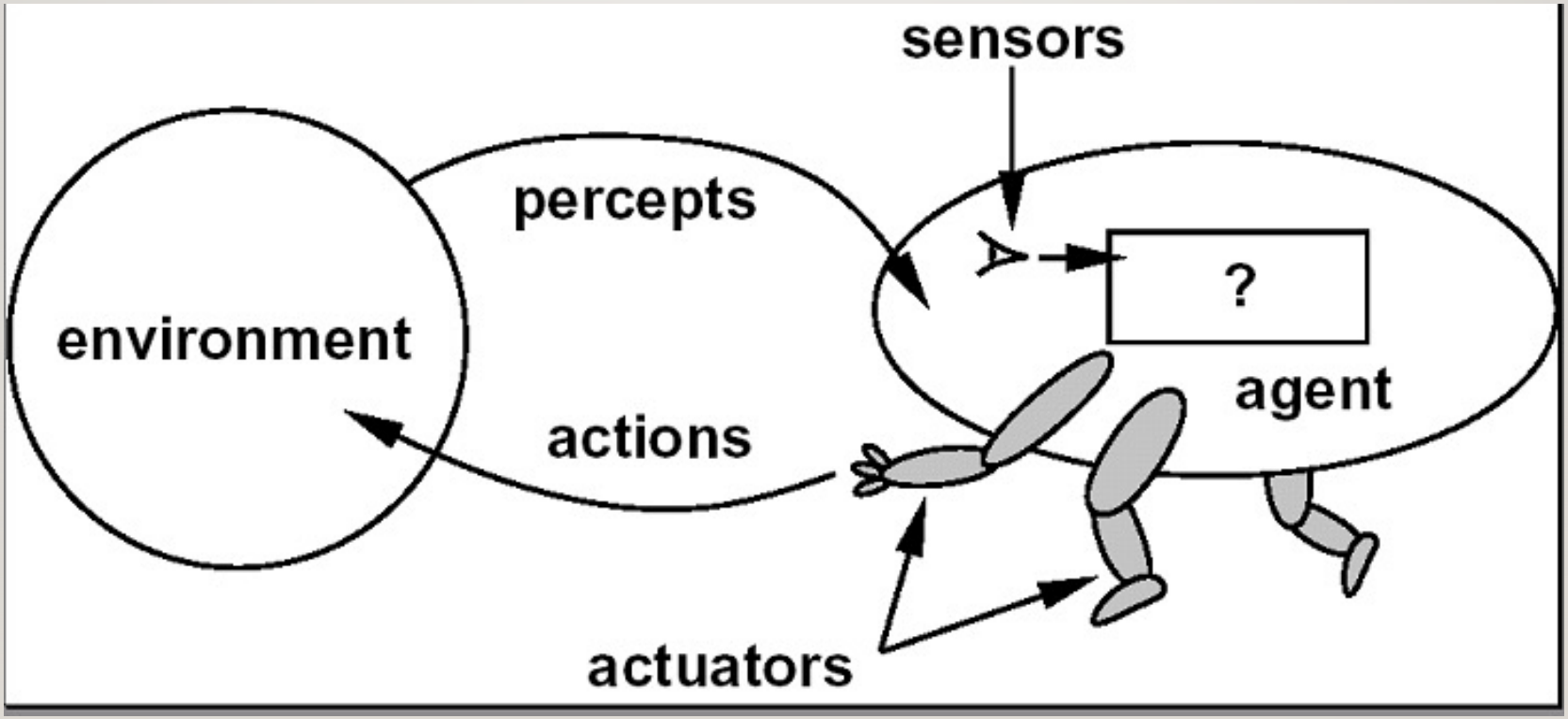
BRIEF HISTORY OF AI

- 1965: Robinson's complete algorithm for logical reasoning
- 1966—74: AI discovers computational complexity; Neural network research almost disappears
- 1969—79: Early development of knowledge-based systems
- 1980—88: Expert systems industry booms
- 1988—93: Expert systems industry busts: `”AI Winter”

BRIEF HISTORY OF AI

- 1985—95: Neural networks return to popularity
- 1988— Resurgence of probability; general increase in technical depth, "Nouvelle AI": ALife, GAs, soft computing
- 1995— Agents…

BACK TO AGENTS



AGENT FUNCTION

- a = F(p)

  where p is the current percept, a is the action carried out, and F is the agent function

- F maps percepts to actions
  $F: P \rightarrow A$

  where P is the set of all percepts, and A is the set of all actions

- In general, an action may depend on all percepts observed so far, not just the current percept, so…

AGENT FUNCTION REFINED

- $a_k = F(p_0\, p_1\, p_2 \ldots p_k)$

  where $p_0\, p_1\, p_2 \ldots p_k$ is the sequence of percepts observed to date, a_k is the resulting action carried out

- F now maps *percept sequences* to actions

  $F: P^* \rightarrow A$

STRUCTURE OF AGENTS

- Agent = architecture + program
- architecture
  - device with sensors and actuators
  - e.g., A robotic car, a camera, a PC, …
- program
  - implements the agent function on the architecture

SPECIFYING THE TASK ENVIRONMENT

- PEAS
- Performance Measure: captures agent's aspiration
- Environment: context, restrictions
- Actuators: indicates what the agent can carry out
- Sensors: indicates what the agent can perceive

PROPERTIES OF ENVIRONMENTS

- Fully versus partially observable
- Deterministic versus stochastic
- Episodic versus sequential
- Static versus dynamic
- Discrete versus continuous
- Single agent versus multiagent

EXAMPLE: MINI CASINO WORLD

- Two slot machines
- Costs 1 peso to play in a machine
  - Takes 10 seconds to play in a machine
- Possible pay-offs: 0, 1, 5, 100
- Given:
  - Amount of money to start with
  - Amount of time to play
  - Expected payoff for each machine
- Objective: end up with as much money as possible

MINI CASINO WORLD

- PEAS description?
- Properties
  - Fully or partially observable?
  - Deterministic or stochastic?
  - Episodic or sequential?
  - Static or dynamic?
  - Discrete or continuous?
  - Single agent or multi-agent?

TYPES OF AGENTS

- Reflex Agent
- Reflex Agent with State
- Goal-based Agent
- Utility-Based Agent

- Learning Agent