First Lab – it is essential to be there!

- Thursday, Jan 14 starting 8:30AM
- We will discuss your ideas about projects
- Each Grad student will have MAX 5 mins to tell us the topic of their research so we can see if there is interest from the class in ‘clustering’ around such kind of work
- Based on the discussions at the Lab we will decide together if we go for final exam or for project.

What did we learn at first class?

- AI is about creating intelligent entities, with a range of abilities such as language, vision, manipulation/navigation..
- Intelligence involves **knowledge** - this must be **represented** and **reasoned** with.
- Solving problems involves **goal-setting** and **planning** (by **reasoning**) on how to reach the goal

AI prehistory (INTERDISCIPLINARY)

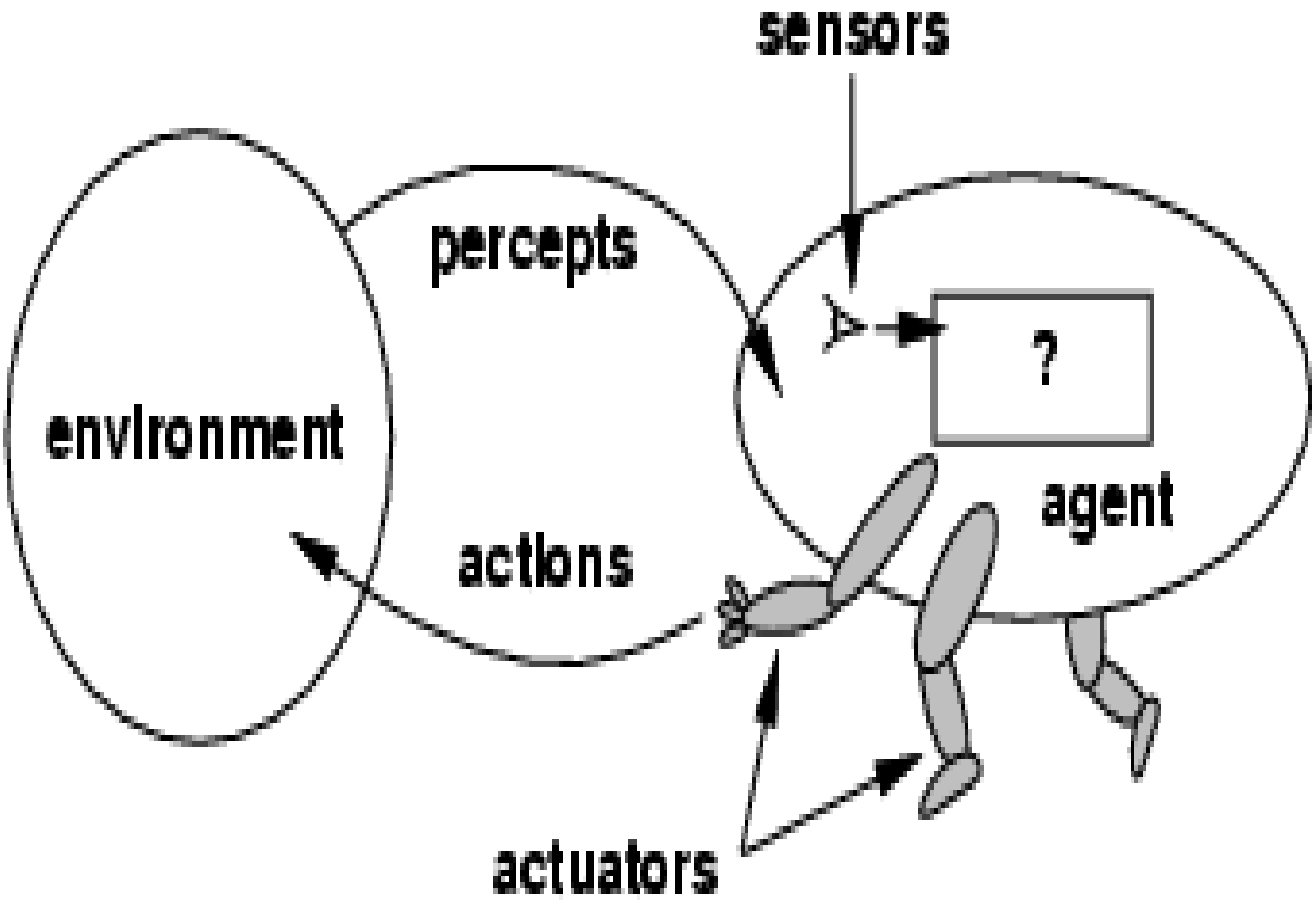
• **Philosophy**	**Logic**, **methods of reasoning**, mind as physical system, foundations of learning, language, rationality
• **Mathematics**	**Formal representation and proof algorithms**, computation, (un)decidability, (in)tractability, probability
• Economics	utility, decision theory
• Neuroscience	physical substrate for mental activity
• Psychology	phenomena of **perception** and motor control, experimental techniques
• **Computer** engineering	building fast computers
• Control theory	design systems that maximize an objective function over time
• Linguistics	**knowledge representation**, grammar

What is AI?

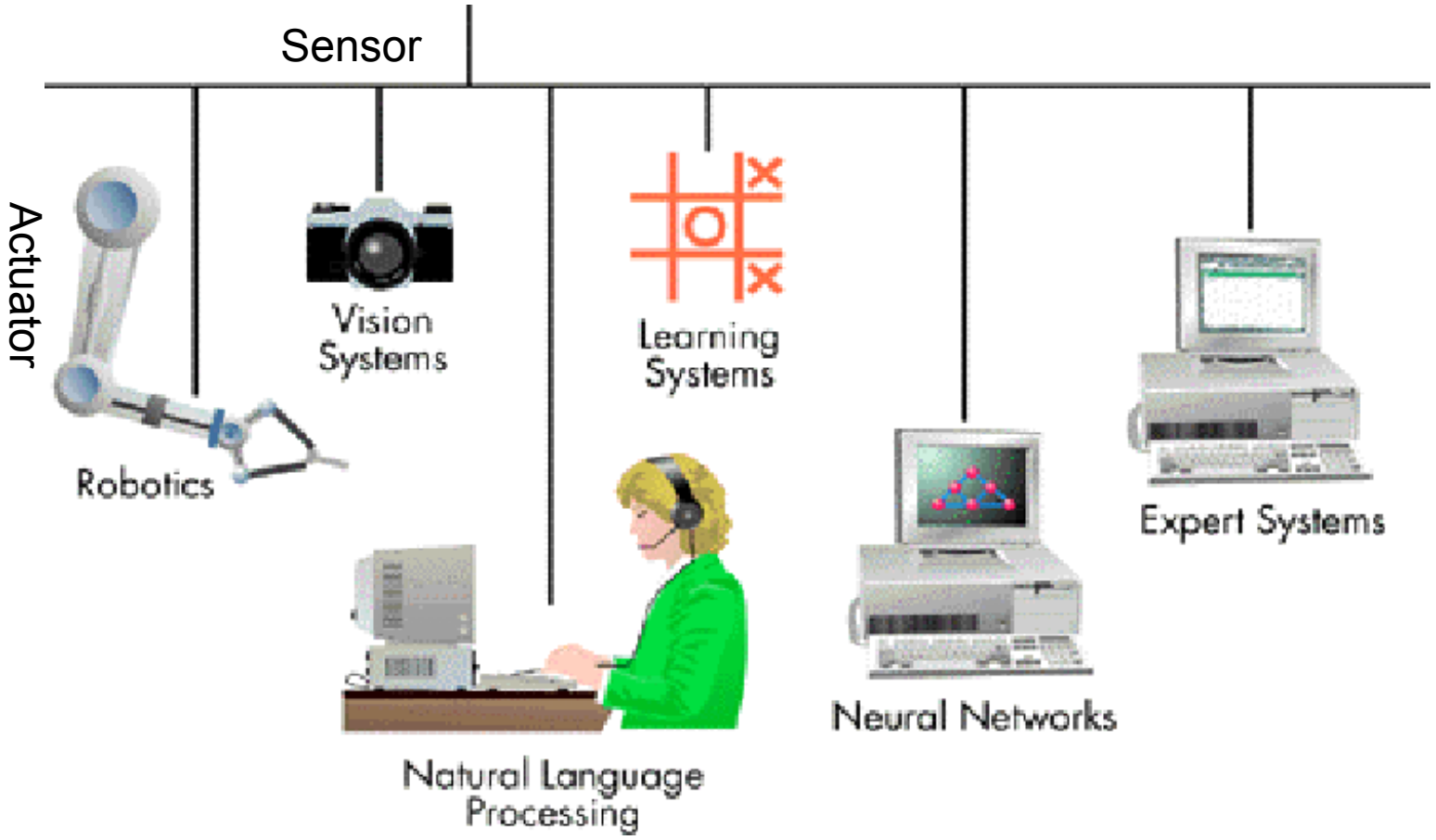
- AI attempts to *build* intelligent entities
- AI is both science and engineering:
  - the *science* of understanding intelligent entities — of developing theories which attempt to explain and predict the nature of such entities;
  - the *engineering* of intelligent entities.

Intelligent Behavior

- **Learn** from experience
- **Apply knowledge** acquired from experience
- Handle complex situations
- **Solve problems when important information is missing**
- **Determine what is important**
- **React quickly and correctly** to a new situation
- Understand visual images
- Process and manipulate symbols
- Be creative and imaginative
- Use heuristics


sensors
percepts
environment
?
agent
actions
actuators

Intelligent Behavior



‘Ingredients’

- Perceptive system
  - A system that approximates the way a human sees, hears, and feels objects
- Vision system
  - Capture, store, and manipulate visual images and pictures
- Robotics
  - Mechanical and computer devices that perform tedious tasks with high precision
- Expert system
  - Stores knowledge and makes inferences

More ‘ingredients’

- Learning system
  - Computer changes how it functions or reacts to situations based on feedback
- Natural language processing
  - Computers understand and react to statements and commands made in a “natural” language, such as English
- Neural network
  - Computer system that can act like or simulate the functioning of the human brain

Intelligent Behavior

Systems that think like humans	**Systems that think rationally**
Systems that act like humans	**Systems that act rationally**

AI as *acting humanly* — as typified by the Turing test

AI as *thinking humanly* — cognitive science.

AI as *thinking rationally* — as typified by logical approaches.

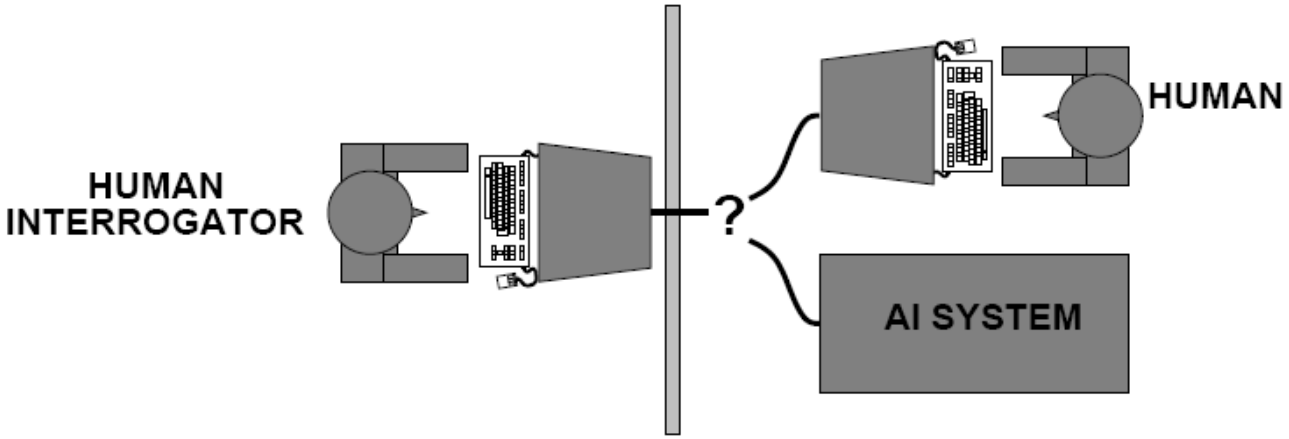
AI as *acting rationally* — the intelligent agent approach.

Acting Humanly

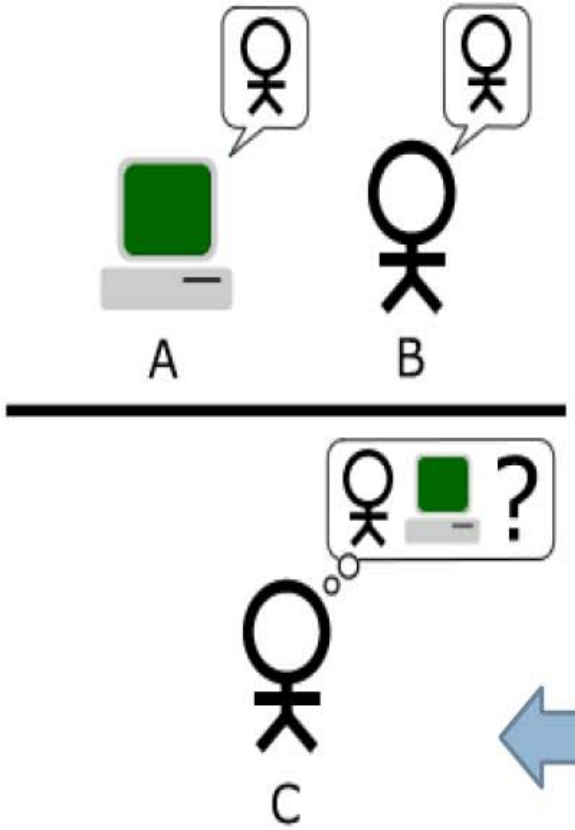
- Emphasis on how to *tell* that a machine is intelligent, not on how to *make it* intelligent
  when can we count a machine as being intelligent?

"Can machines think?" ⟶ "Can machines behave intelligently?"

- Most famous response due to Alan Turing, British mathematician and computing pioneer:



- Can a machine be intelligent?
  - Turing test: Test if a machine is intelligent



If he/she cannot tell who is human and who is machine, then the machine is “intelligent”.

Turing Test

- No program has yet passed Turing test!
  (Annual Loebner competition & prize.)
- A program that succeeded would need to be capable of:
  - natural language understanding & generation;
  - knowledge representation;
  - learning;
  - automated reasoning.
- Note no *visual* or *aural* component to basic Turing test — augmented test involves video & audio feed to entity.

Problem: Turing test is not *reproducible*, *constructive*, or amenable to *mathematical analysis*

Example conversation with system winning Loebner prize

- P: No worries! Hey! I didn't come here to be subjected to bad language! How rude!
- J: Sorry, sorry. Didn't mean to offend. So, how old are you?
- P: That's OK. Just don't do it again. Well, should my age really matter. Just for the record I'm 24. How old are you then. But.. System was based on fairly simple methods, with pre-stored amusing replies.
- These simple methods first used in "Eliza" - a program to emulatepsychotherapist.

(http://www.manifestation.com/neurotoys/eliza.php3).

- **Human-like performance doesn't guarantee intelligence.**
- **Lets discuss this at the Lab…**

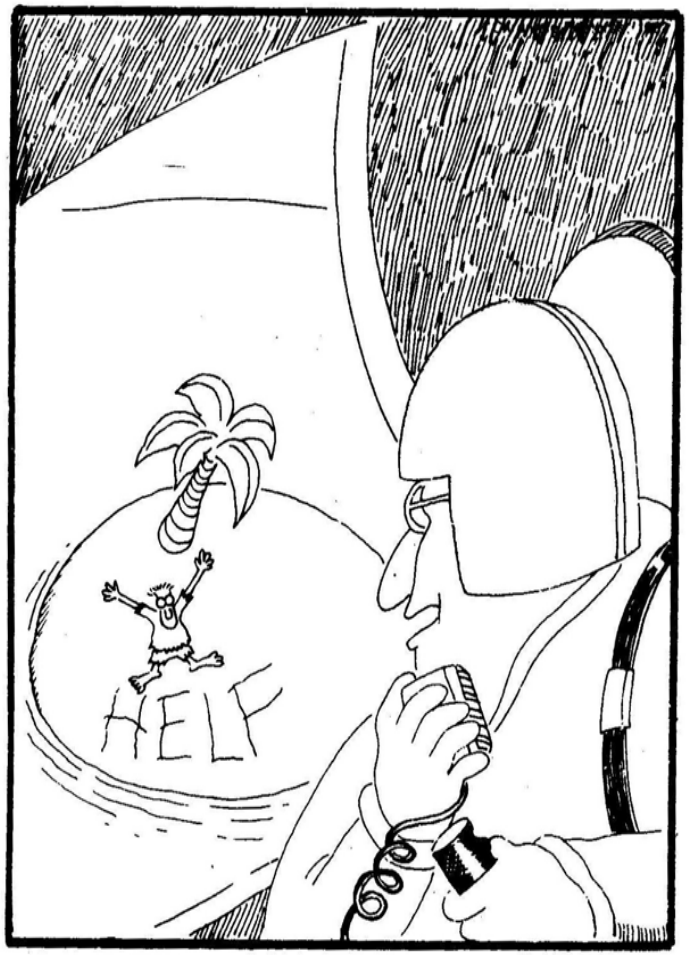
What is AI?

Views of AI fall into four categories:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

'**Emulating the Mind**' (see my website, main page, second paragraph)

Embodied agents



"Wait! Wait! . . . Cancel that, I guess it says 'helf.'"

Emotions and AI

- Growing awareness: role of
  - Environment
  - Resource constraints
- Emotions encapsulate relevant information
  - Contribute to intelligent behavior
  - Support “social” interaction

Computational models of emotions

- Ethology-inspired models
  - *Model agents, environment, survival constraints*
- Emotion-related learning
  - *Provide feedback (reinforcement) on behaviors to guide learning*
- Appraisal-based models
  - *Features of situation map to emotional states*
- Architecture-level models
  - *Combination of mechanisms to support successful behavior (e.g. Sloman's CogAff)*

What is AI?

Views of AI fall into four categories:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

Thinking Humanly

- Try to understand how the mind works — how do we think?
- Two possible routes to find answers:
  - by *introspection* — we figure it out ourselves!
  - by *experiment* — draw upon techniques of psychology to conduct controlled experiments. ("Rat in a box"!)
- The discipline of *cognitive science*: particularly influential in *vision*, *natural language processing*, and *learning*.

Thinking humanly: cognitive modeling

- 1960s "cognitive revolution": information-processing psychology
- Requires scientific theories of internal activities of the brain
- -- How to validate? Requires

  1) Predicting and testing behavior of human subjects (top-down)

  or 2) Direct identification from neurological data (bottom-up)
- Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

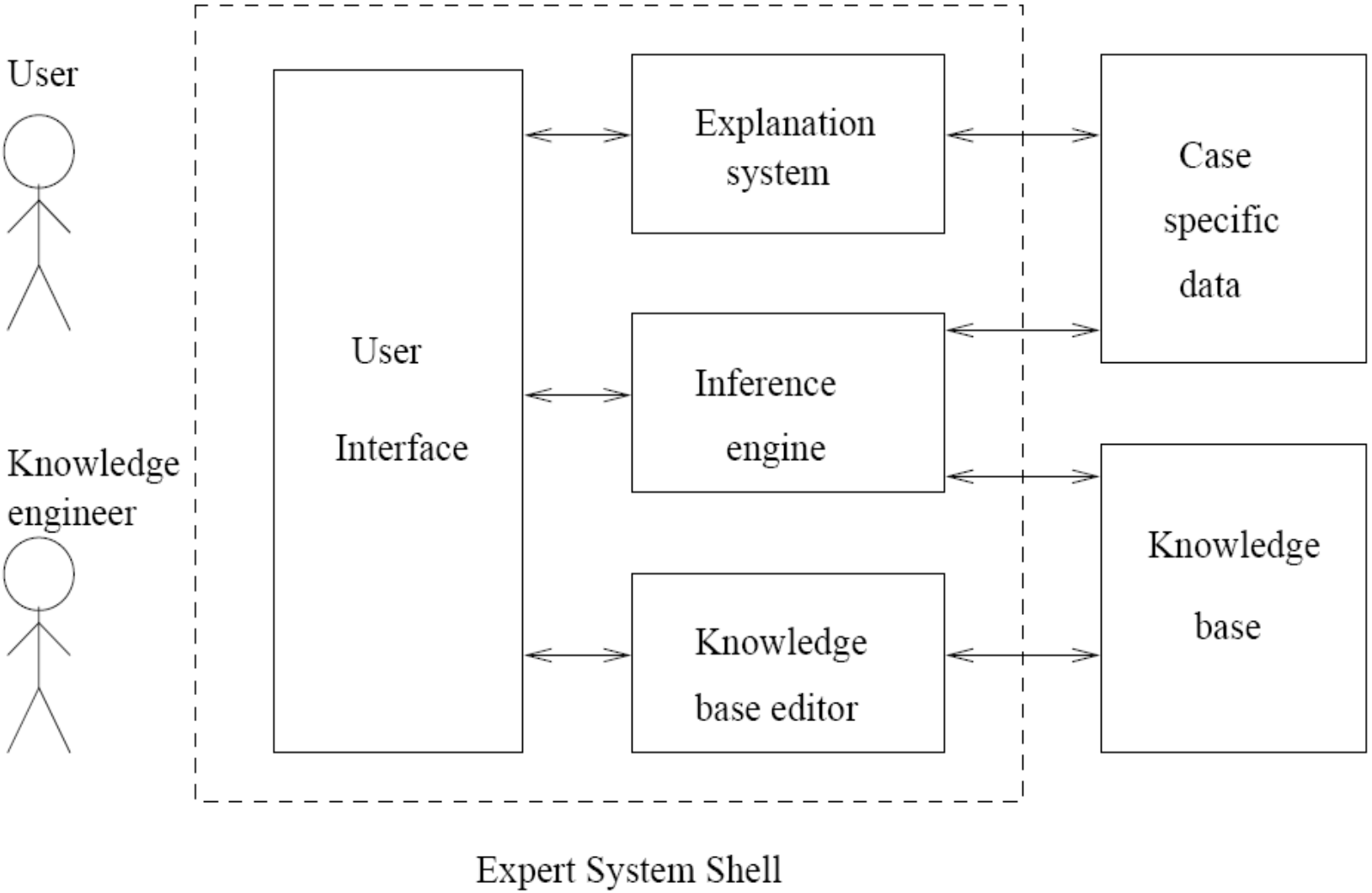
Human vs. Machine Thinking

Expert systems — “AI success story in early 80’s”

- Human expert’s knowledge and experience is passed to a computer program
- Rule-based representation of knowledge
- Typical domains are:
  - medicine (INTERNIST, MYCIN, . . .)
  - geology (PROSPECTOR)
  - chemical analysis (DENDRAL)
  - configuration of computers (R1)

Thinking humanly works

Architecture of an Expert System I



Rule-based Systems

- Knowledge is specified as a collection of *rules*.
- Each rule has the form

$$condition \longrightarrow action$$

  which may be read if $condition$ then $action$.
- The $condition$ (antecedent) is a *pattern*.
- The *action* (consequent) is an operation to be performed if rule *fires*.
- Knowledge is applied to *facts*—unconditional statements that are assumed to be correct (at the time they are used).

What is AI?

Views of AI fall into four categories:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

Thinking rationally: "laws of thought"

- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of *logic*: *notation* and *rules of derivation* for thoughts; may or may not have proceeded to the idea of mechanization
- Direct line through mathematics and philosophy to modern AI
- Problems:
  1. Not all intelligent behavior is mediated by logical deliberation
  2. What is the purpose of thinking? What thoughts should I have?

Thinking Rationally

- Trying to understand how we *actually* think is one route to AI — but how about how we *should* think.
- Use *logic* to capture the *laws of rational thought* as *symbols*.
- *Reasoning* involves shifting symbols according to well-defined rules (like algebra).
- Result is *idealised* reasoning.

What is AI?

Views of AI fall into four categories:

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

The textbook advocates "acting rationally"

Acting rationally: rational agent

- Rational behavior: ***doing the right thing***
- The right thing: that which is expected to ***maximize goal achievement***, given the available information
- Doesn't necessarily involve thinking – e.g., blinking reflex – but thinking should be in the service of rational action

Designing Rational Agents

- An agent is an entity that perceives and acts
- Abstractly, an agent is a function from percept histories to actions:

$$[f: \mathcal{P}^{\star} \rightarrow \mathcal{A}]$$

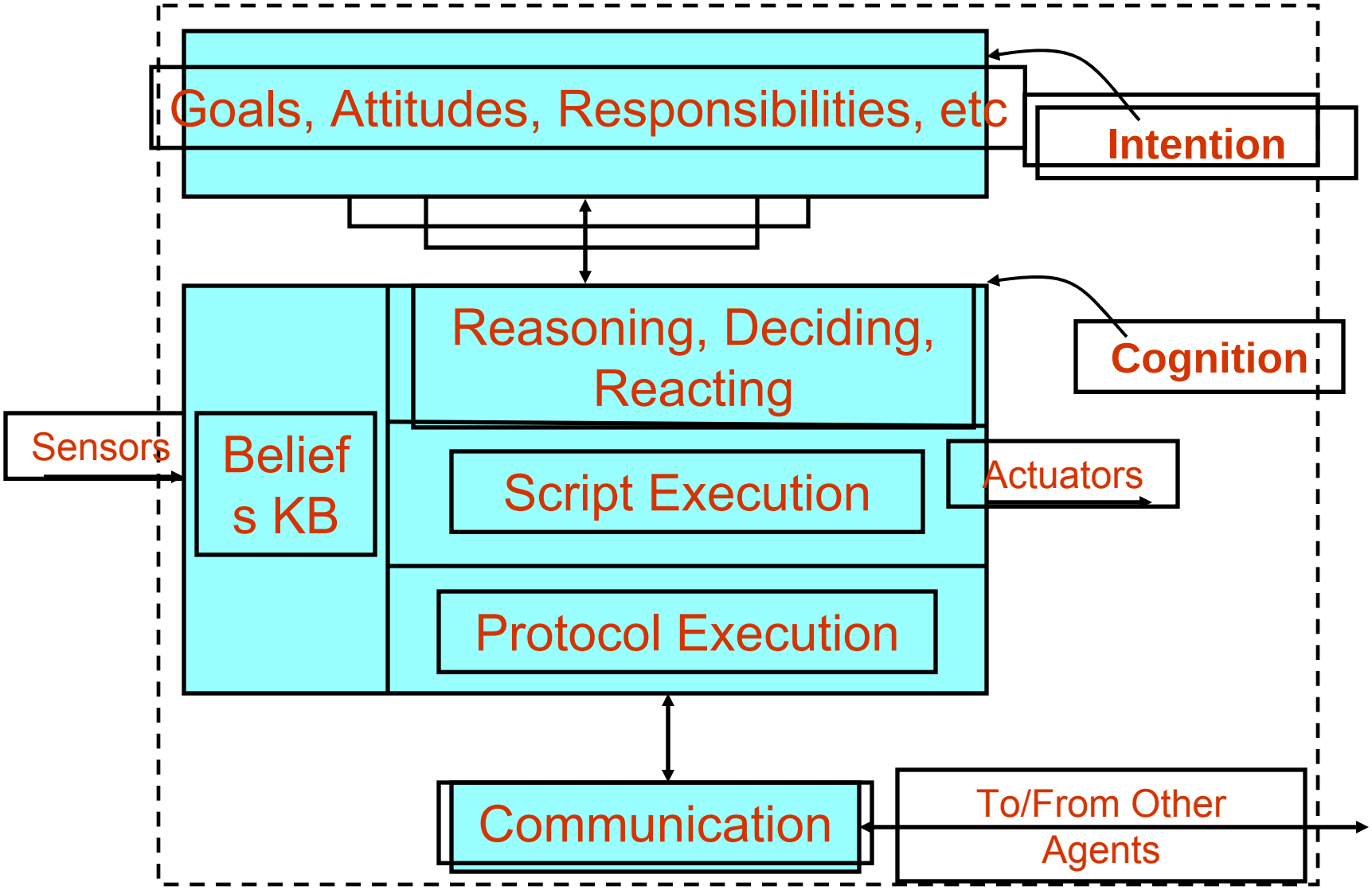
- For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance
- Caveat: computational limitations make perfect rationality unachievable

  → design best program for given machine resources

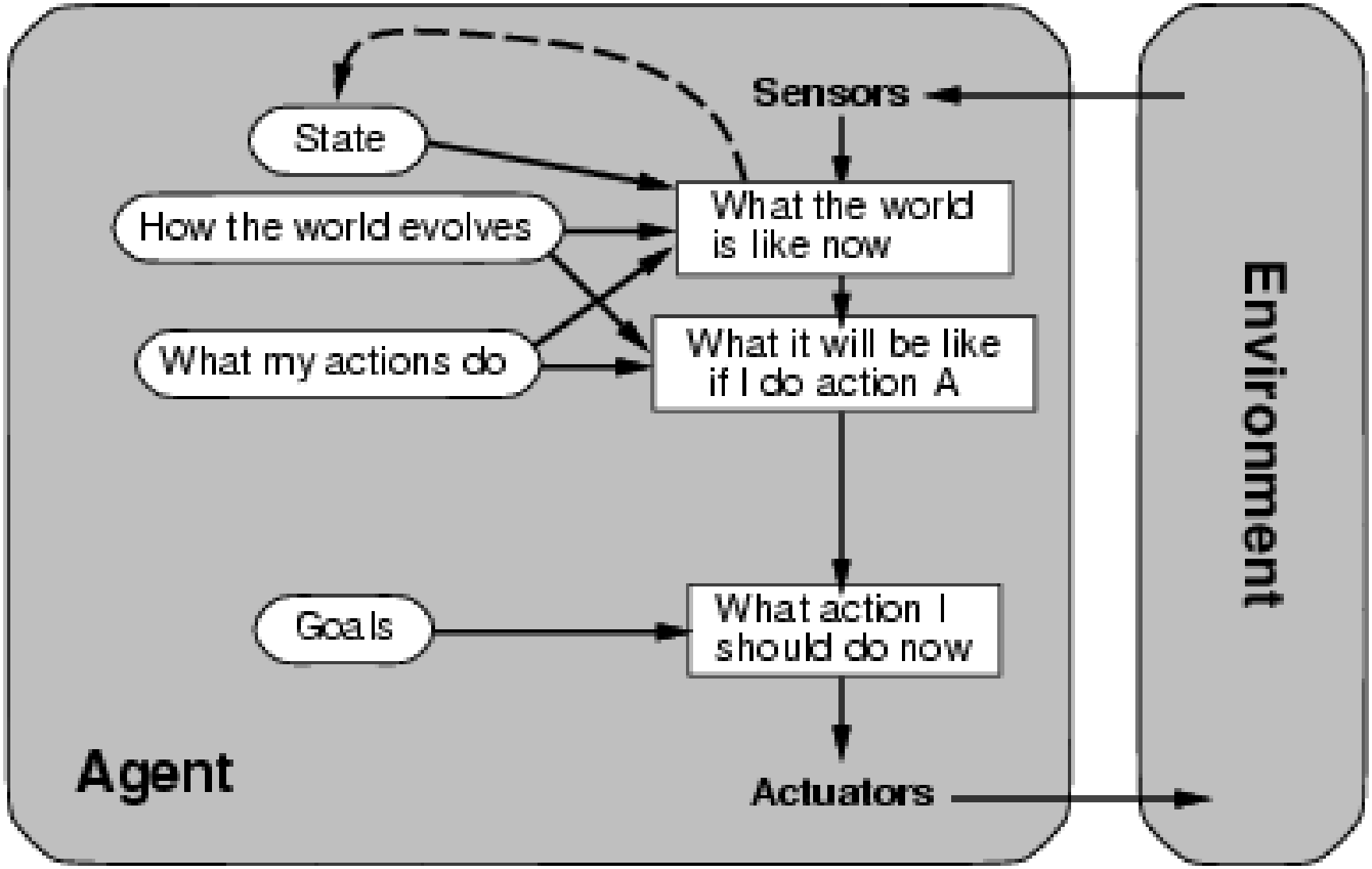
Acting Rationally

- Acting rationally = acting to achieve one's goals, given one's beliefs.
- An *agent* is a system that perceives and acts; intelligent agent is one that acts rationally w.r.t. the goals we delegate to it.
- Emphasis shifts from designing *theoretically best* decision making procedure to best decision making procedure possible in *circumstances*.
- Logic may be used in the service of finding the best action — not an end in itself.

Human (or advanced robot) agent



Goal-based agents



ASSIGNMENTS

- 1. Read Chapter 1 in your book
- 2. **(Due Jan. 14 when we meet at the lab)**

Pick your favorite question at the end of Chapter 1 and write a min one phrase, max one page essay with your answer. Reflect on what is of utmost interest to you (e.g. relative to your quest in your [graduate] studies, or your personal / hobby interest) from the field of AI – and how can you use this class to dive into this particular field.