

Assignment 1

Thomas Bakken Moe

1. Define artificial intelligence (AI). Find at least 3 definitions of AI that are not covered in the lecture.

Three definitions I've found:

- "The art of creating machines that perform functions that require intelligence when performed by people."
- "The study of the computations that make it possible to perceive, reason, and act."
- "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, learning ..."

2. What is the Turing test, and how [is it] conducted?

The Turing test is a test designed by Alan Turing in 1950 designed to test a machine's ability to exhibit behaviour that is indistinguishable from that of a human.

The test is conducted by having a human (the interrogator) conduct a 5-minute text message conversation with either a human or the program we want to test. After the 5 minutes are up, the interrogator will guess if who they've been chatting with is a real human or a program pretending to be human. If the program manages to fool the interrogators 30% of time, it will have passed the Turing test.

3. What is the relationship between thinking rationally and acting rationally? Is rational thinking an absolute condition for acting rationally?

Thinking rationally and acting rationally often go together, but are not absolutely bound together. Thinking rationally is trying to reduce problems to a perfect mathematical representation (logic) and finding a perfect solutions to them. Rational acting is where this thinking hits the real world. In the real world, the agent's information will not be perfect and thus the problem cannot be reduced down and solved perfectly mathematically.

Acting rationally is not acting out the perfect solution, but rather the best solution possible given imperfect information.

So no, rational thinking is not an absolute condition for acting rationally.

4. Describe rationality. How is it defined?

The Merriam-Webster dictionary defines rationality as "The quality or state of being rational". Rational is defined as "Having reason or understanding". Thus, a valid definition of rationality is: "The quality or state of having reason or understanding."

5. What is Aristotle's argument about the connection between knowledge and action? Does he make any further suggestion that could be used to implement his idea in AI? Who was/were the first AI researcher(s) to implement these ideas? What is the name of the program/system they developed? Google about this system and write a short description about it.

Aristotle argues that the connection between knowledge and action is reasoning. He argues that good humans are rational beings and will use knowledge and reasoning to perform rational actions to achieve their goals. Aristotle's ideas relate pretty directly to goal-based agents. One of the first implementations of this goal-based approach was Shakey the robot. Shakey the robot was designed as a general-purpose robot. An operator could type a generic command like "Push the block off the platform" and Shakey would have to decode the command and interpret how it related to its environment. It would have to locate the platform with the block on top and figure out a way to push the block off.

6. Consider a robot whose task it is to cross the road. Its action portfolio looks like this: look-back, lookforward, look-left-look-right, go-forward, go-back, go-left and go-right.

(a) While crossing the road, a helicopter falls down on the robot and smashes it. Is the robot rational?

In this instance, the agent is rational. Nothing in its action portfolio would allow it to detect the helicopter (unless it throws a cartoon shadow on the ground) and prevent the helicopter from hitting itself. The agent cannot look up. The agent can only act to achieve its goal based upon the information it can perceive. It cannot perceive the helicopter.

(b) While crossing the road on a green light, a passing car crashes into the robot, preventing it from crossing. Is the robot rational?

In this instance, the agent is not rational. The agent's goal is to reach the other side of the road, and its action portfolio gives it the tools needed to detect and avoid the incoming car. Thus, the agent is behaving irrationally. This assumes that the agent is nimble enough to get out of the way of the car.

7. Consider the vacuum cleaner world described in Chapter 2.1 of the textbook. Let us modify this vacuum environment so that the agent is penalized 1 point for each movement.

NOTE: I'm assuming here that the agent does not want to lose points and that it gains points when it sucks dust. Its main goal remains: to clean the entire world.

(a) Can a simple reflex agent be rational for this environment? Explain your answer

A simple reflex agent cannot be rational for this environment. If the agent is on tile A, and tile B is dusty: the agent needs to clean tile B to accomplish its goal. But all the agent knows is that it will get penalized for moving, and thus it cannot risk moving on the chance that B is dusty.

(b) Can a reflex agent with state be rational in this environment? Explain your answer.

This agent is not rational. The fact that the agent has a memory (internal state) does not really matter in my mind. The agent will still get stuck, unable to complete its goal. If the agent is on tile A, and tile B is dusty: the agent needs to clean tile B to accomplish its goal. But all the agent knows is that it will get penalized for moving, and thus it cannot risk moving on the chance that B is dusty.

(c) Assume now that the simple reflex agent (i.e., no internal state) can perceive the clean/dirty status of both locations at the same time. Can this agent be rational? Explain your answer. In case it can be rational, design the agent function.

In this case, the agent will have a perfect observation of the state of the environment (both square A and B). This gives the agent the ability to decide if moving left or right will help it achieve its goal. Now: if the

agent is on square A, and tile B is dusty: the agent will observe that square B is dirty and it knows that cleaning the square will negate the penalty it takes from moving.

Agent function:

Observe environment -> if current square is dirty: clean current tile. Else if other square is dirty: move to other square. Else (both squares are clean) do nothing. -> repeat

8. Consider the vacuum cleaner environment shown in Figure 2.2 in the textbook. Describe the environment using properties from Chapter 2.3.2, e.g. episodic/sequential, deterministic/stochastic etc. Explain selected values for properties in regards to the vacuum cleaner environment.

These are the descriptors I would use for this environment:

- Partially observable
 - The environment can be observed by the agent, but the agent cannot observe the entire state of the environment at the same time. The agent can only know what's happening at the square it is currently inhabiting.
- Single agent
 - There is only one agent acting on the environment.
- Deterministic
 - The way an action from the agent alters the environment is known and always results in the same change. If the agent sucks a dirty square, the square will always be cleaned.
- Sequential
 - This environment is sequential. The actions taken by the agent in the previous time-step affect the state in the current time-step.
- Static
 - The agent is the only force acting upon the environment. There is no chance that a change can occur while the agent is deliberating or performing an action.
- Discrete
 - As presented, the actions that the agent can perform are handled in discrete steps. If the agent moves to the left, nothing else will happen before the agent has completed the movement. For the agent, it is as if there is no time in between moving left and starting the next action.
- Known
 - The agent knows all its possible actions and how they affect the environment. It knows that moving to the left will always result in the agent ending up in the square to the left of it.

9. Discuss the advantages and limitations of these four basic kinds of agents:

(a) Simple reflex agents

Simple reflex agents are simple to implement and require few resources. They are however of limited maximal intelligence. Their intelligence might be perfectly adequate for relatively simple problems. For more complex problems however: they might be found lacking.

(b) Model-based reflex agents

Model-based reflex agents differ from simple reflex agents by having an "internal state" (e.g. a memory). This allows these agents to get a better understanding of the environment. They can for example interpret a car's speed by comparing its position "now" and its position from the last time-step. The types of

problems this kind of agent can solve are a lot more than what simple reflex agents can solve. They do however require slightly more resources than their simpler counterpart.

(c) Goal-based agents

Reflex agents react directly to input from the environment and respond with a "pre-planned" response. Goal-based will reason how their actions will result in them getting closer to their goal. Goal-based agents are capable of handling much more complex environments than simple reflex agents. They are however less efficient than what a reflex agent would be. They also have no way of determining the best way of reaching their goal. All the agent cares about is reaching the goal, no matter how inefficient.

(d) Utility-based agents

Utility-based agents does not work to strictly reach their goal. They work to maximize their utility. The way they "earn utility" is to achieve progress towards their goal. The better the progress, the more utility. The idea is that shorter/less costly routes to their goal gives more utility. These agents can achieve impressive intelligence, but they are complex to implement and can costly.