Lecture 1: Introduction

Frederik Mallmann-Trenn

(version 1.3)

Today

1. Introduction to Artificial Intelligence
2. Probabilistic Reasoning I
3. Probabilistic Reasoning II
4. Sequential Decision Making
5. Temporal Probabilistic Reasoning
6. Game Theory
7. Argumentation I
8. Argumentation II
9. (A peek at) Machine Learning
10. AI & Ethics

Intelligent agents

- There are many different definitions of AI
- We're going to take an agent-based perspective on artificial intelligence.
- Because:
  - One of the main ways that AI is viewed at the moment.
  - Main way that we think about AI.

Intelligent agents

- Complementary to module on Agents and Multi-Agent Systems (6CCS3AMS/7CCSMAMS).

- Focus here is on decision-making in complex environments.

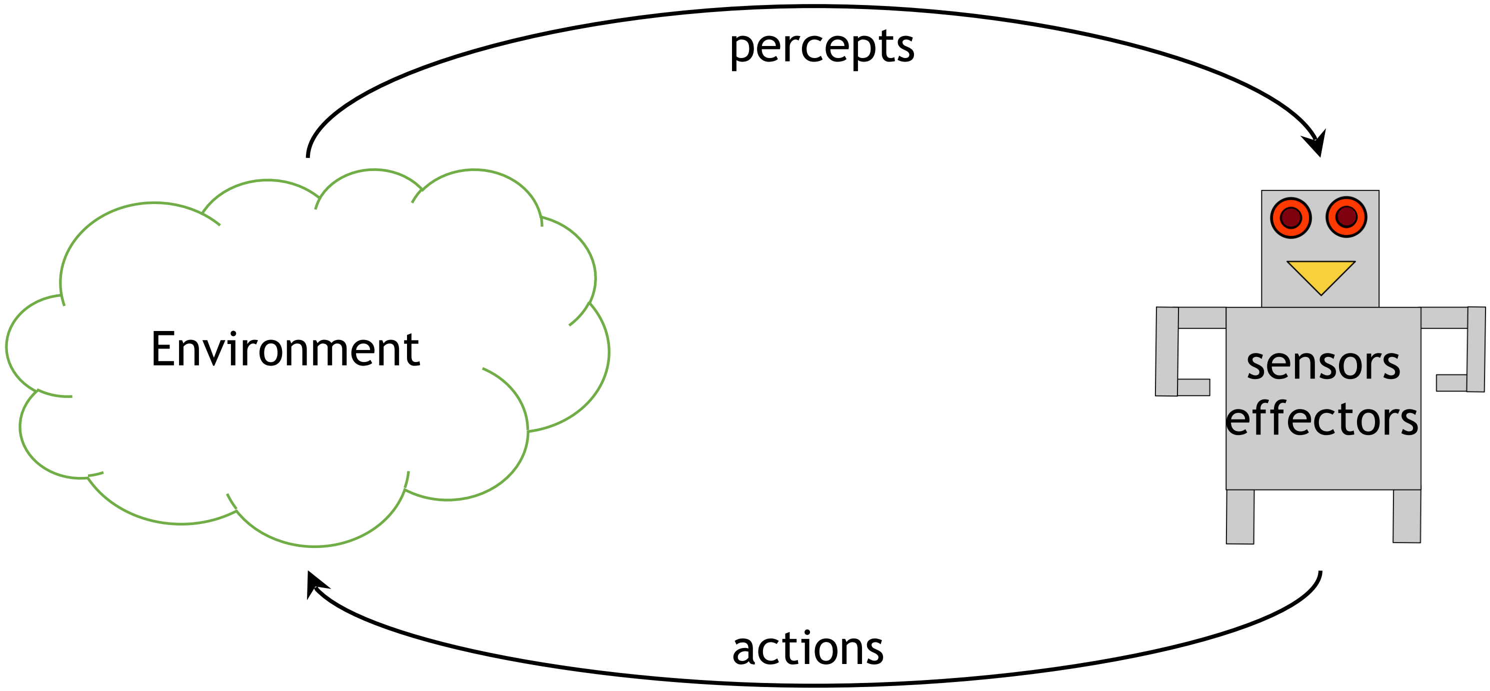
- Focus there is on mechanisms to allow agents to operate effectively in a multi-agent environment.

Intelligent agents

- An agent is a system that:
  - is situated in an environment,
  - is capable of perceiving its environment, and
  - is capable of acting in its environment

  with the goal of satisfying its design objectives.

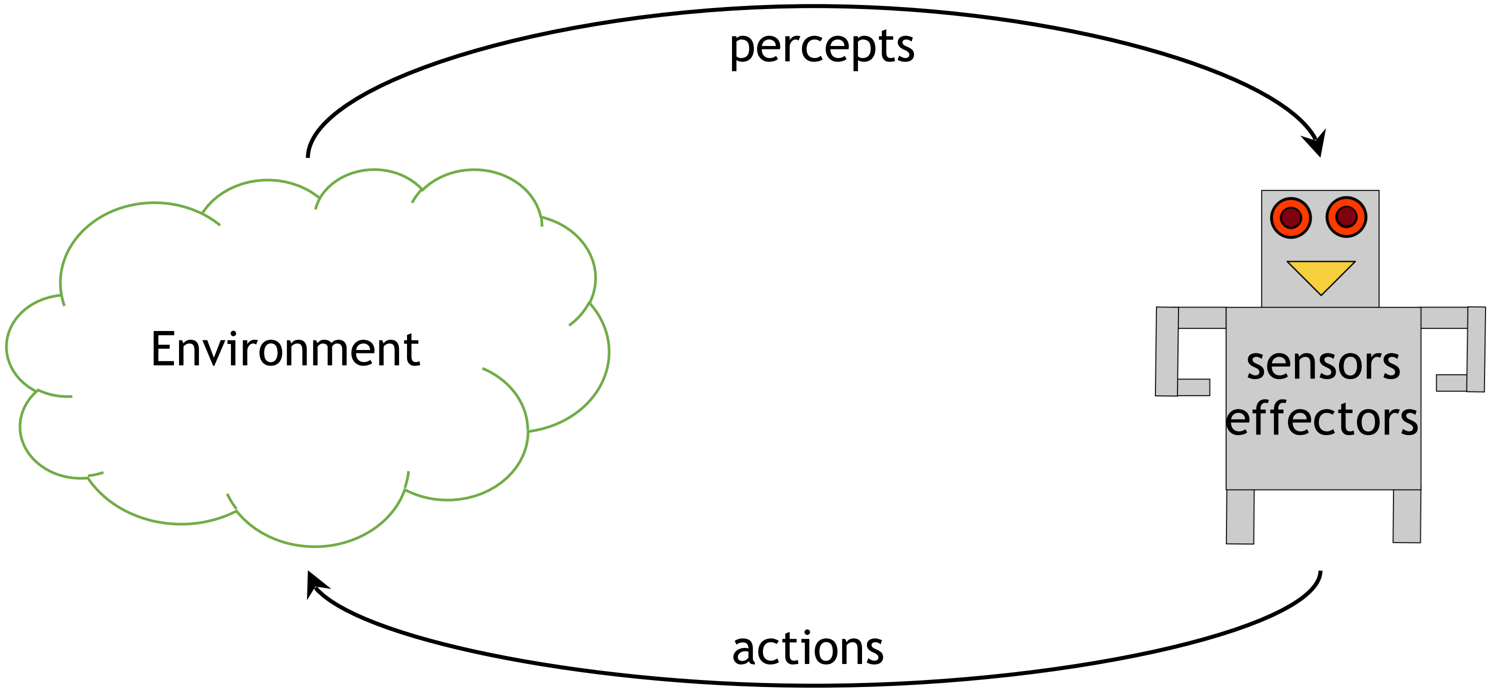
Intelligent agents



- An agent is a system that:
  - is situated in an environment,
  - is capable of perceiving its environment, and
  - is capable of acting in its environment

  with the goal of satisfying its design objectives.

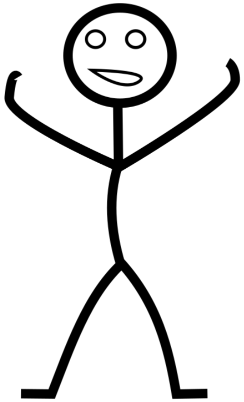
Intelligent agents



Task is to program the agent to convert percepts to actions.

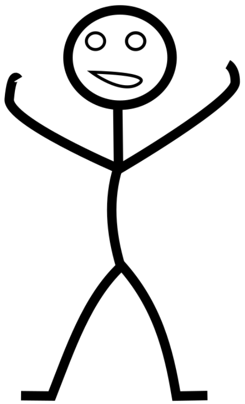
Intelligent agents

- Human "agent":
  - environment:
  - sensors:
  - effectors:



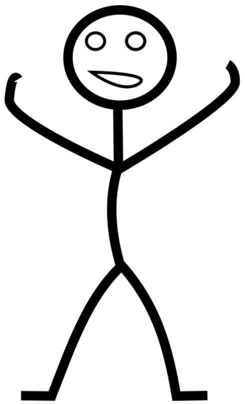
Intelligent agents

- Human "agent":
  - environment: physical world;
  - sensors:
  - effectors:



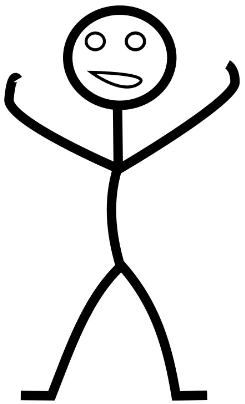
Intelligent agents

- Human “agent”:
  - environment: physical world;
  - sensors: eyes, ears, tastebuds, ...
  - effectors:



Intelligent agents

- Human “agent”:
  - environment: physical world;
  - sensors: eyes, ears, tastebuds, ….
  - effectors: hands, legs, voice, ….



Intelligent agents

- Software "agent" (e.g. personal digital assistants like Alexa, Siri):
  - environment:
  - sensors:
  - effectors:

Credit: Masaki

Intelligent agents

- Software "agent" (e.g. personal digital assistants like Alexa, Siri):
  - environment: internet and physical world;
  - sensors:
  - effectors:

Credit: Masaki

Intelligent agents

- Software “agent” (e.g. personal digital assistants like Alexa, Siri):
  - environment: internet and physical world;
  - sensors: microphone, gather data from internet sources (e.g. weather, traffic,…), access calendar…
  - effectors:

Credit: Masaki

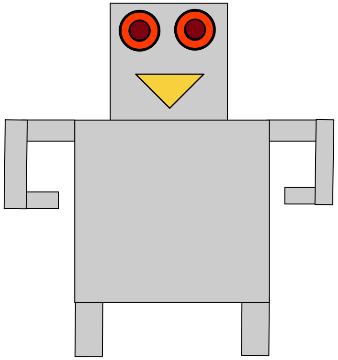
Intelligent agents

- Software “agent” (e.g. personal digital assistants like Alexa, Siri):
  - environment: internet and physical world;
  - sensors: microphone, gather data from internet sources (e.g. weather, traffic,…), access calendar…
  - effectors: sound, play music, book ubers, order things online, arrange appointments…

Credit: Masaki

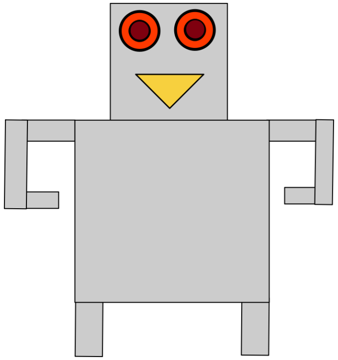
Intelligent agents

- Robot "agent":
  - environment:
  - sensors:
  - effectors:



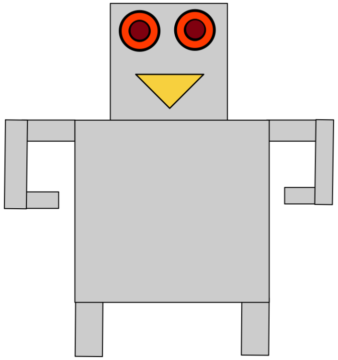
Intelligent agents

- Robot "agent":
  - environment: physical world;
  - sensors:
  - effectors:



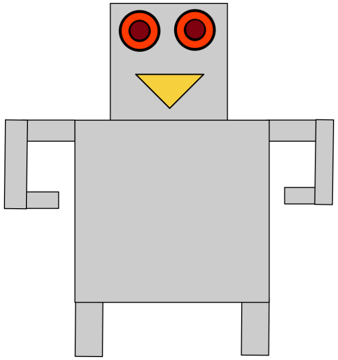
Intelligent agents

- Robot "agent":
  - environment: physical world;
  - sensors: sonar, camera, microphone, pressure sensors …
  - effectors:



Intelligent agents

- Robot “agent”:
  - environment: physical world;
  - sensors: sonar, camera, microphone, pressure sensors …
  - effectors: screen, sound, wheels, arm, gripper, sucker….



Personal digital assistants

Credit: Masaki
Tokutomi

Robot companions

Credit: Xavier Caré

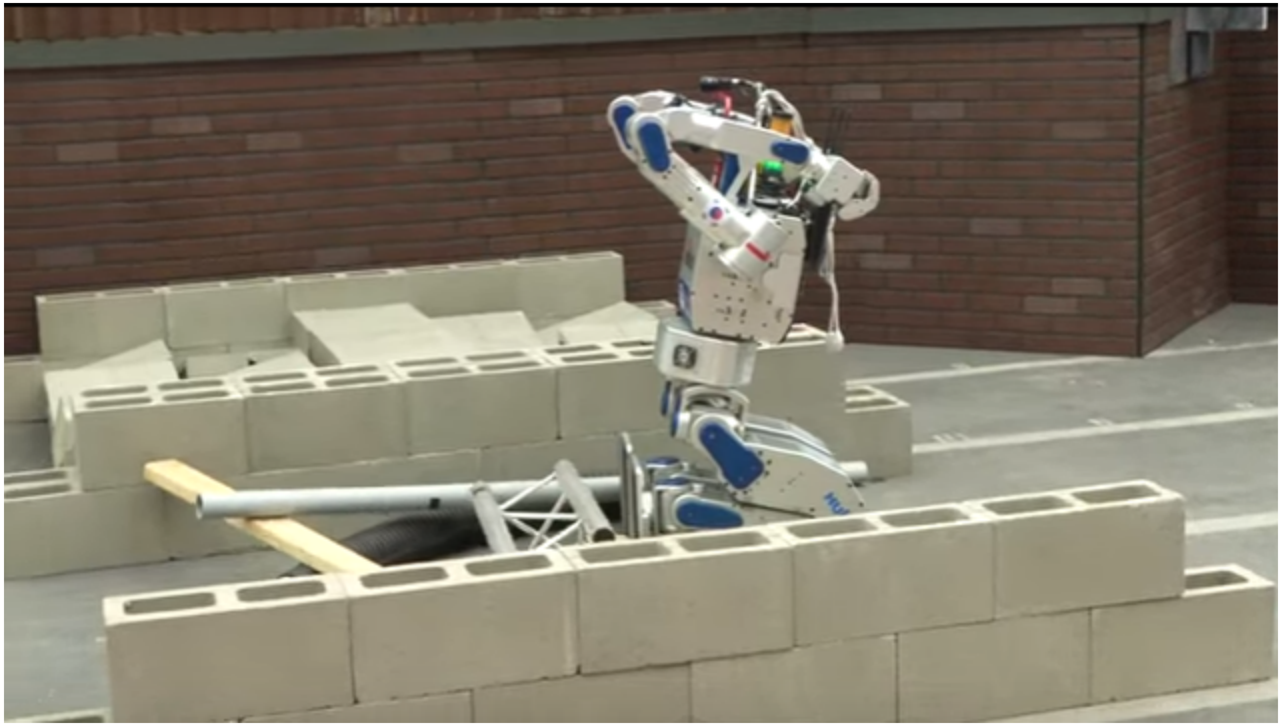
Robot companions

Credit: Kim Kyung Hoon / Reuters

Mars rover

Credit: NASA/JPL-Caltech/MSSS

Disaster response robots



Credit: Darpa

Darpa Robotics Challenge Program

https://www.youtube.com/watch?v=mpsXQCHrAlM

The dark side ...

BBC | Sign in | News | Sport | Weather | iPlayer | Sounds

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Family & Education

Technology

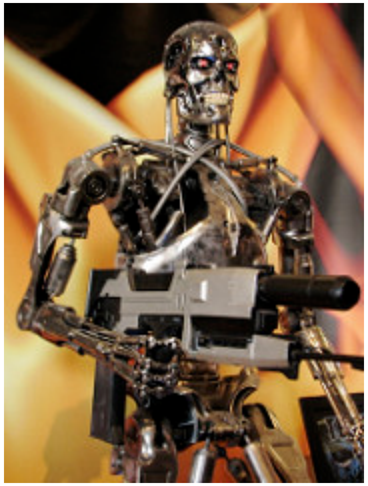
Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones
Technology correspondent

2 December 2014

Share

What about the ethics of AI?

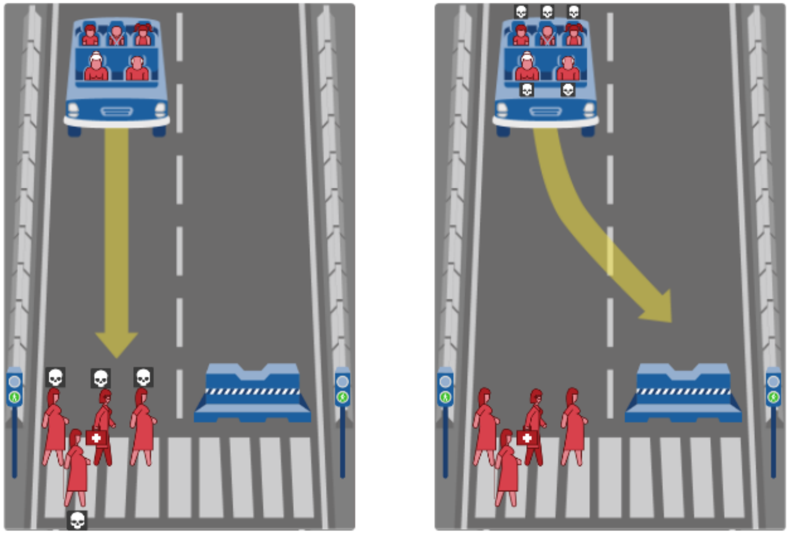


Credit: Roo Reynolds, Flikr

http://medicalfuturist.com/will-robots-take-over-our-jobs-in-healthcare/

Amazon Robotics

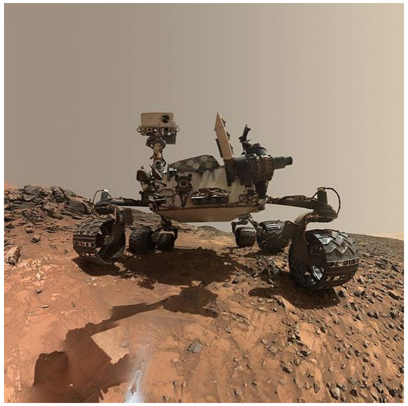
U.S. Air Force photo/Lt Col Leslie Pratt - USAF photo via public domain website http://www.af.mil/shared/media/photodb/photos/081131-F-7734Q-001.jpg



http://moralmachine.mit.edu/

What should be the goal? What is the right thing?

- Back to our robots. Say you want to build a Mars rover, how do you program it to explore Mars?



- We don’t always know what we should do.

- Knowing what to do can in principle be easy:
  - Consider all the alternatives and choose the “best”.

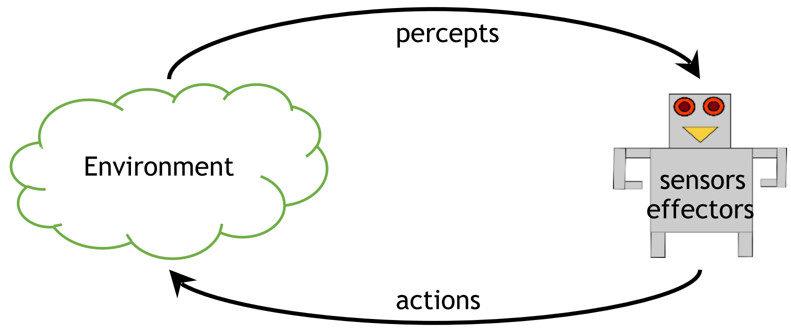
- <u>Class activity</u>:
  - Why isn’t this easy in practice?
  - Give examples where it’s not easy to decide what the right thing to do is.

Do the right thing?

- Why isn’t this easy in practice?

- Need to be able to:
  - identify the alternatives,
  - decide which alternative has the best outcome,
  - and do so quickly enough for the decision to be useful!

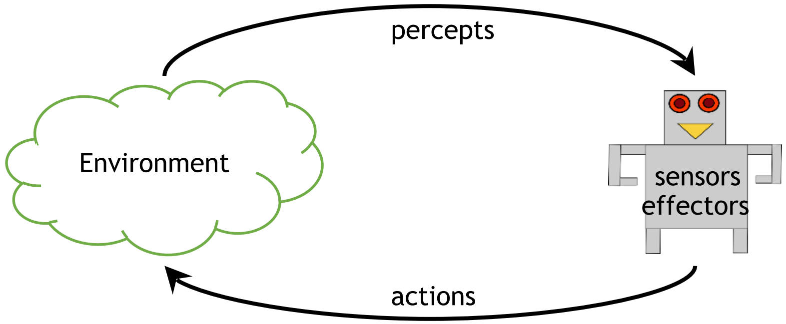
Ideal rational agent



For each percept sequence, an ideal rational agent will act to maximise its performance measure, on the basis of information provided by the percept sequence plus any information built into the agent.

Note this does not preclude performing actions to find things out.

Ideal rational agent



- Rational agent doesn't mean omniscient or clairvoyant agent.
- Rational agent doesn't mean perfect agent.

I'm walking along the Strand and see an old friend. There's no traffic near by and the green man is showing, I don't have anywhere else I need to be, so being rational I start to cross the road. Meanwhile, high above, a cargo door falls off a passing airliner, and before I reach my friend I am flattened. Was I irrational to cross the street?

Sequence of percepts

Sequence of percepts

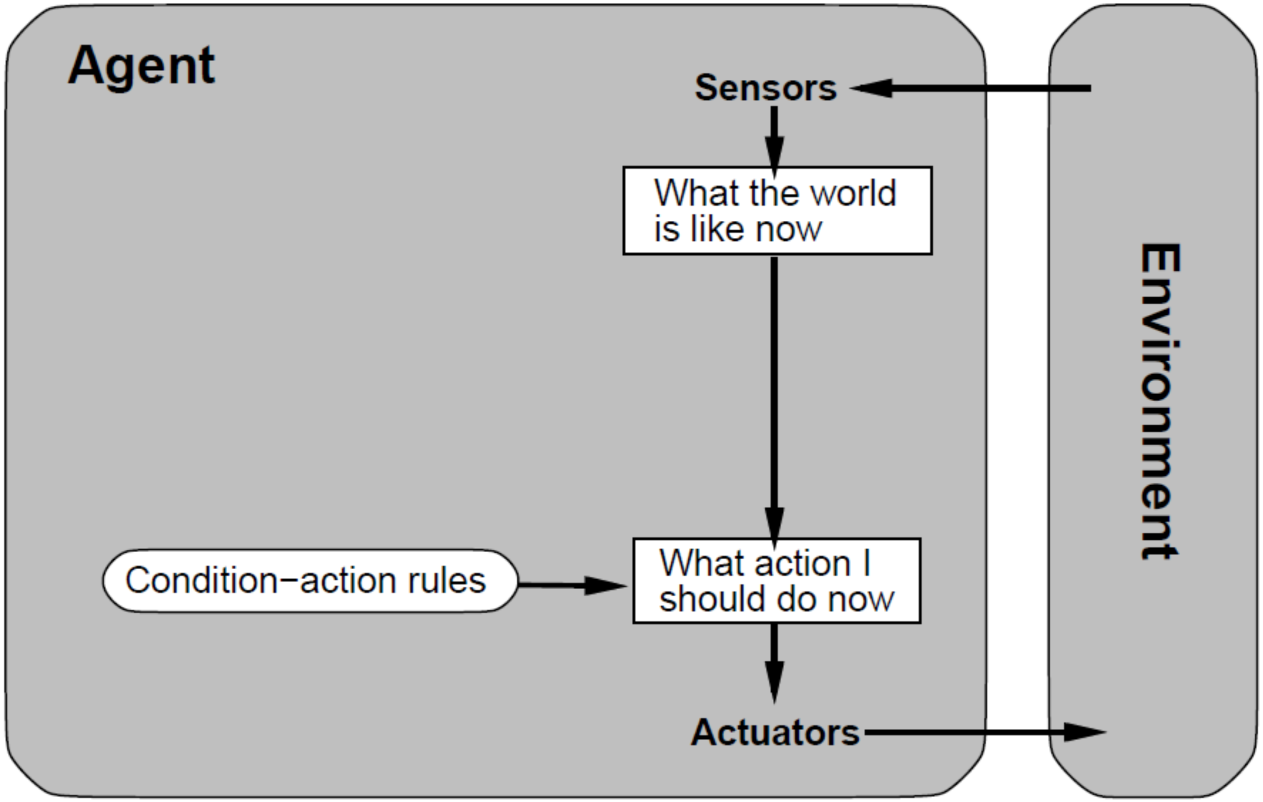
Class activity: What are some of the challenges for an agent in dealing with percepts like this?

Sequence of percepts

Some challenges:

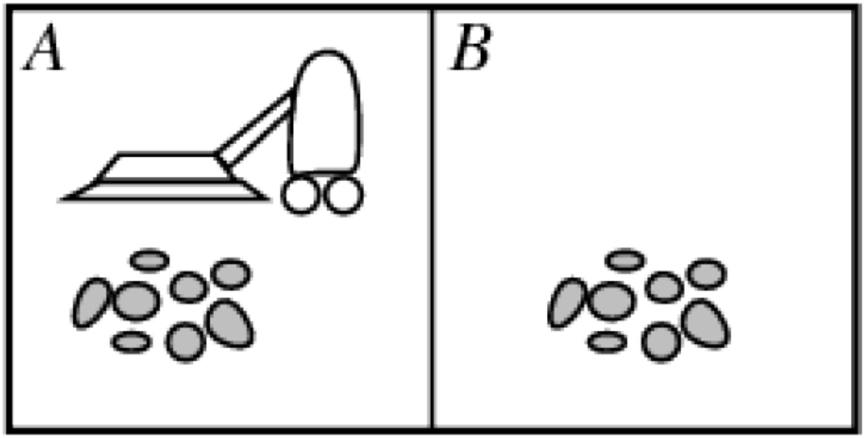
- what to represent and how to represent it,
- interpreting the percepts.

Simple reflex action



A simple agent maps each percept directly to an action.

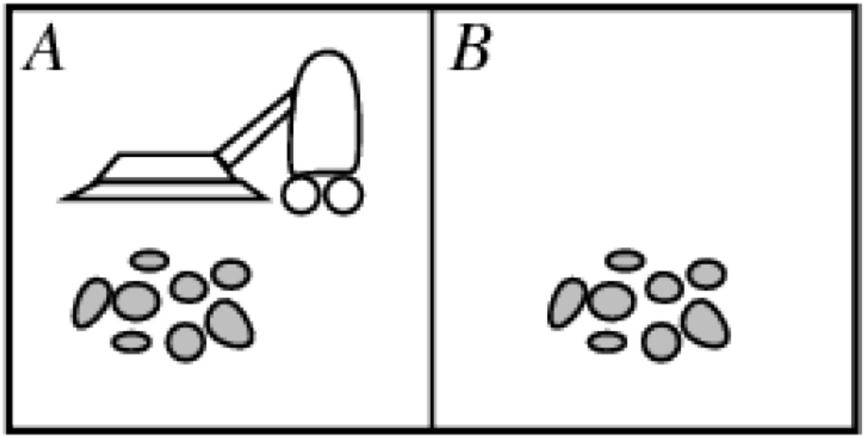
Vacuum world



- Sensors:
  - Location? (A or B)
  - Dirty? (Yes or No)
- Actions:
  - Suck
  - Right
  - Left

Class activity: Come up with a prioritised list of **if** *condition* **then** *action* rules to control a simple vacuum agent.

Vacuum world

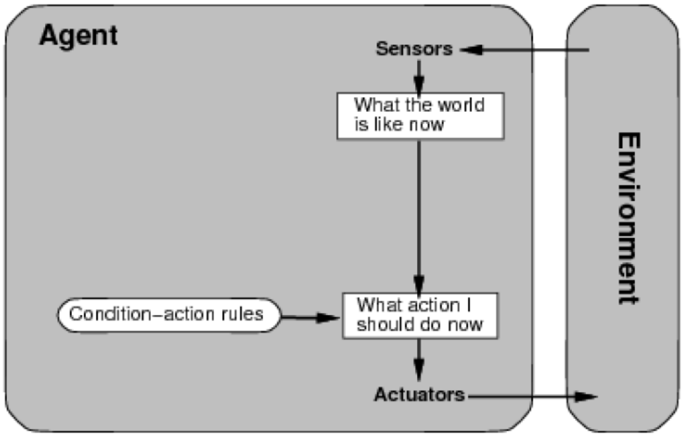


- Sensors:
  - Location? (A or B)
  - Dirty? (Yes or No)

- Actions:
  - Suck
  - Right
  - Left

if *Dirty = Yes* **then** *Suck*
else if *Location = A* **then** *Right*
else if *Location = B* **then** *Left*

Simple reflex agent

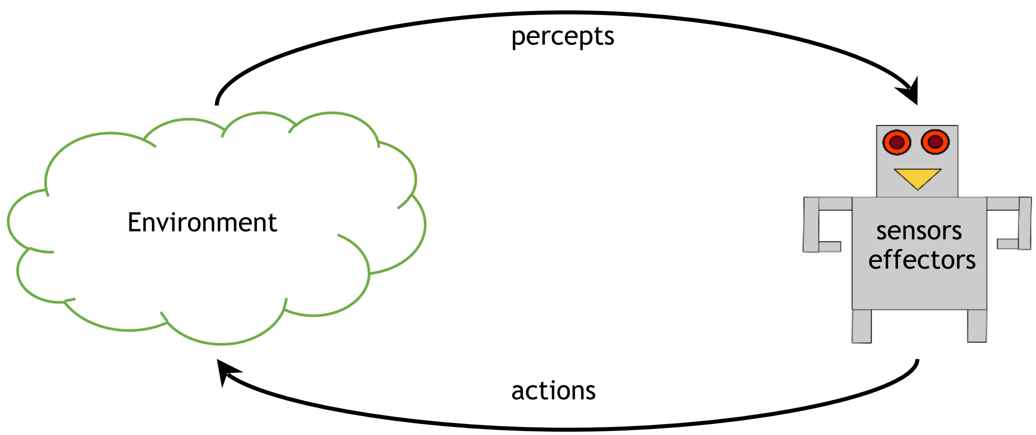


function Simple-Reflex-Agent(*percept*) **returns** an action

static: *rules*, a set of condition-action rules

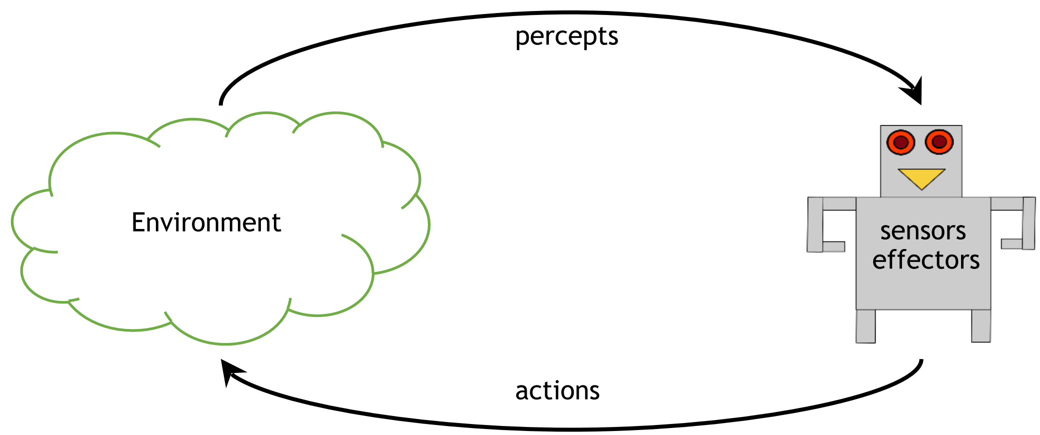
state ← Interpret-Input(*percept*)
rule ← *Rule-Match*(*state*, *rules*)
action ← *rule*.*action*
return *action*

Most of the time things are not so simple



Most of the time things are not so simple

<u>Class activity</u>: What features of a complex environment can make it difficult to decide what the right thing to do is?



Most of the time things are not so simple

Can the agent detect all relevant aspects of the state?

An environment is *accessible* if the agent can obtain all accurate and up-to-date information about all the relevant aspects.

Most of the time things are not so simple

Can the agent detect all relevant aspects of the state?

An environment is *accessible* if the agent can obtain all accurate and up-to-date information about all the relevant aspects.

E.g., chess vs Starcraft

Most of the time things are not so simple

Can the agent be certain what state will result from its actions?

A *deterministic* environment is one in which any action has a single guaranteed effect – there's no uncertainty about the state that will result from performing an action.

A *stochastic* environment is one in which we can quantify the non-determinism with probability.

Most of the time things are not so simple

Can the agent be certain what state will result from its actions?

A *deterministic* environment is one in which any action has a single guaranteed effect – there's no uncertainty about the state that will result from performing an action.

A *stochastic* environment is one in which we can quantify the non-determinism with probability.

E.g., chess vs roulette

Most of the time things are not so simple

Can the agent’s decision about what to do affect future decisions?

In a *sequential* environment, the current decisions affect the future decisions

In an *episodic* environment, the agent's experience is divided into atomic "episodes" (each episode consists of the agent perceiving and then performing a single action) and the choice of action in each episode depends only on the episode itself.

Most of the time things are not so simple

Can the agent’s decision about what to do affect future decisions?

In a *sequential* environment, the current decisions affect the future decisions

In an *episodic* environment, the agent's experience is divided into atomic "episodes" (each episode consists of the agent perceiving and then performing a single action) and the choice of action in each episode depends only on the episode itself.

E.g., chess vs expert advise systems

Most of the time things are not so simple

Can the agent be sure that its environment will only change as a result of its actions?

A *static* environment remains unchanged except by the performance of the agent's actions.

A *dynamic* environment has other processes operating on it, and hence changes in ways beyond agent's control.

E.g., Sudoku vs chess

Most of the time things are not so simple

Does the agent’s environment have a fixed, finite number of actions and percepts in it?

If yes, then it’s discrete, otherwise it’s continuous.

E.g., walk ‘north’ or ‘south’ vs turn 42.2328891… degrees.

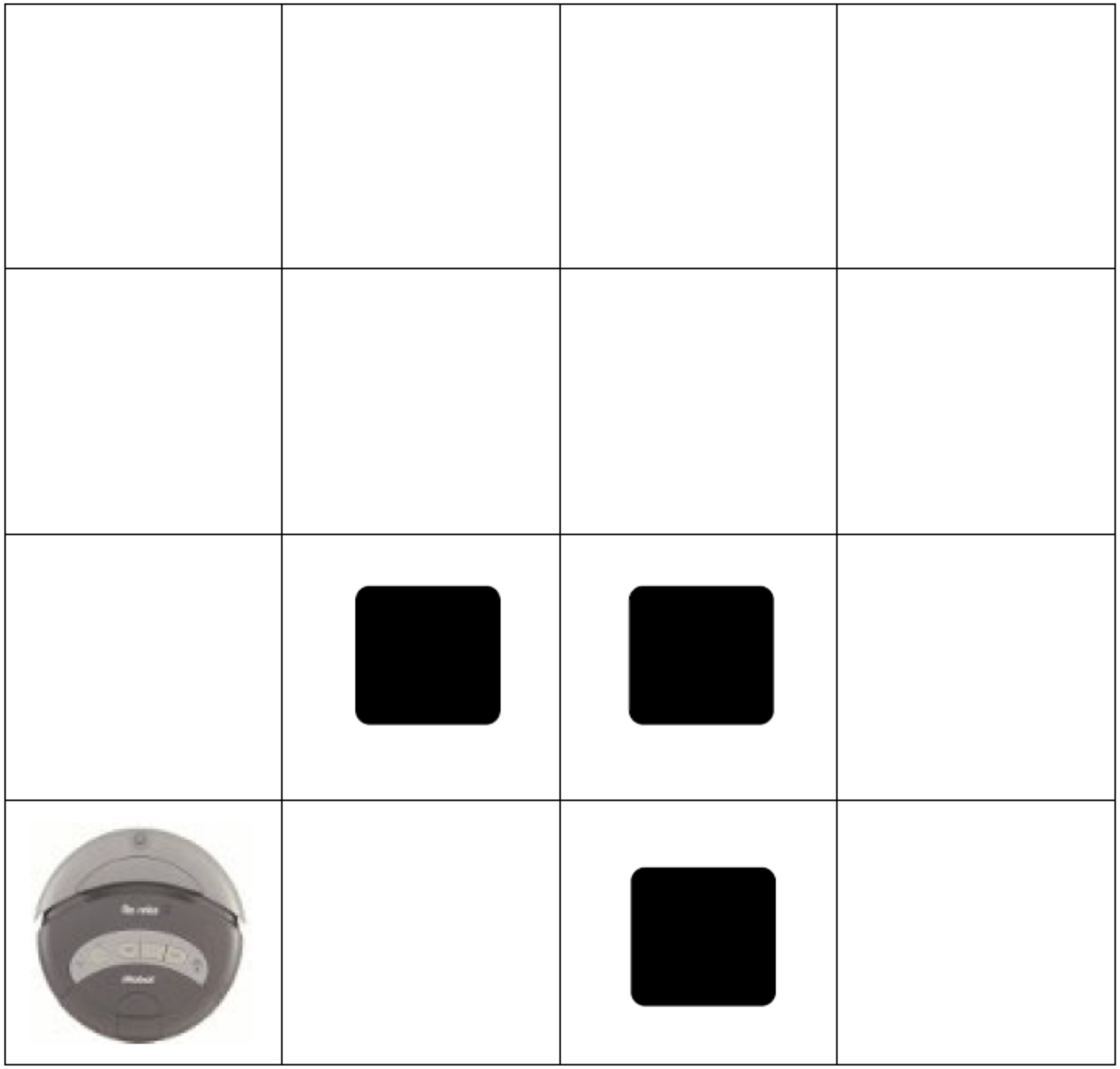
A run of an agent in an environment

When an agent acts in an environment it generates a run.

A *run* is a sequence of interleaved states and actions

s0, a1, s1, a2, s2, a3, s3 ….

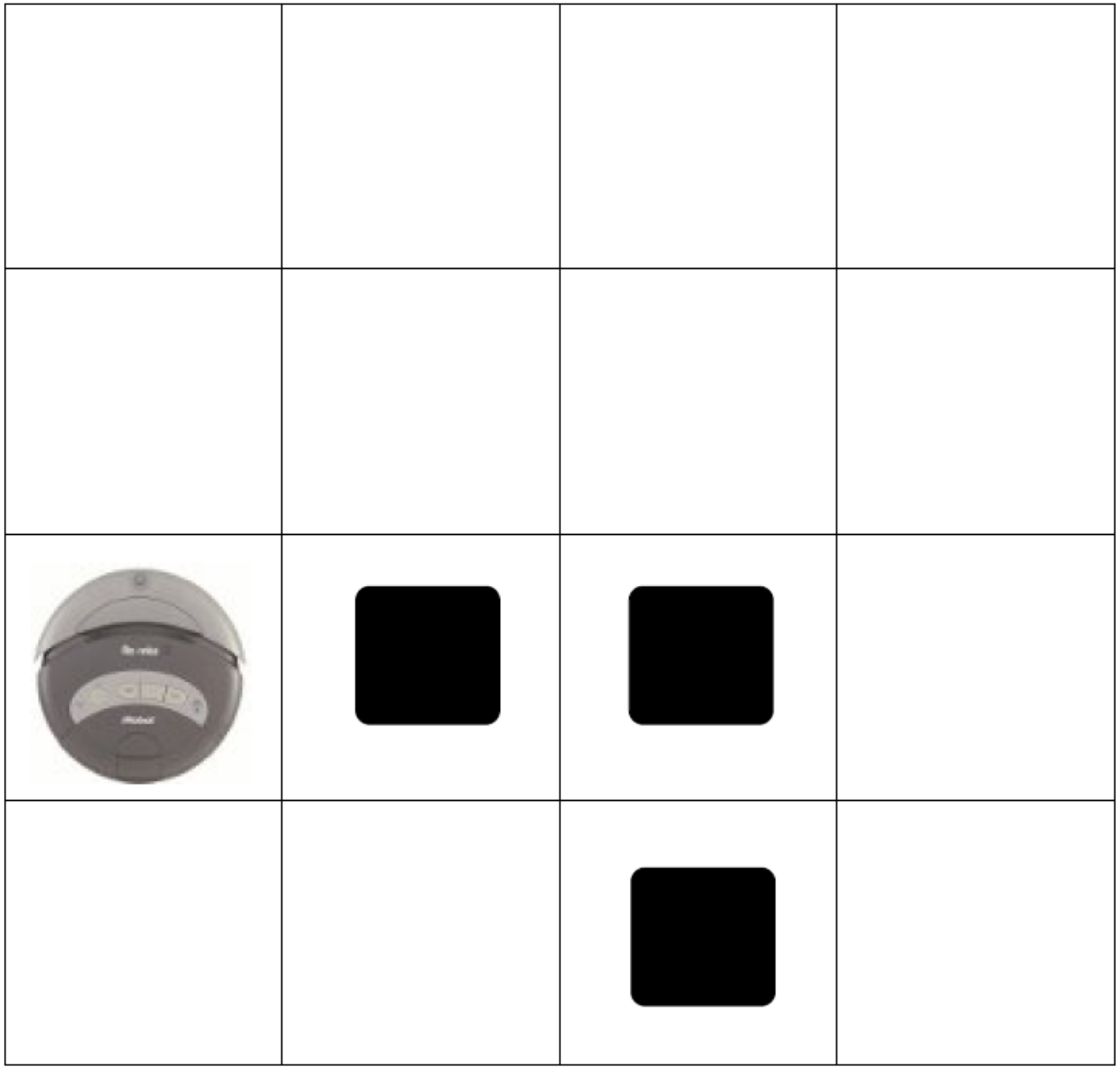
A run of an agent in an environment



A run of an agent in an environment

North!

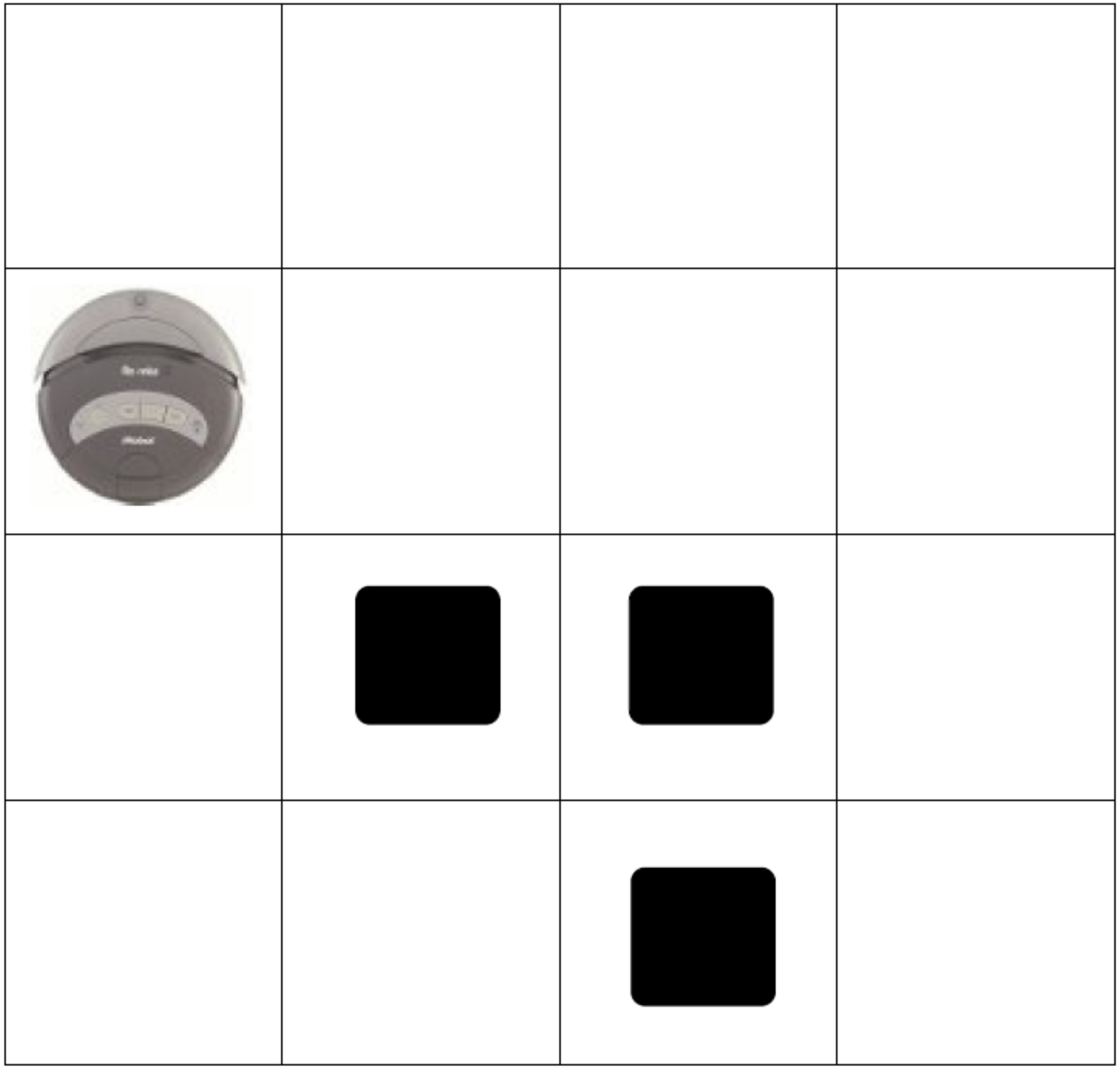
A run of an agent in an environment



A run of an agent in an environment

North!

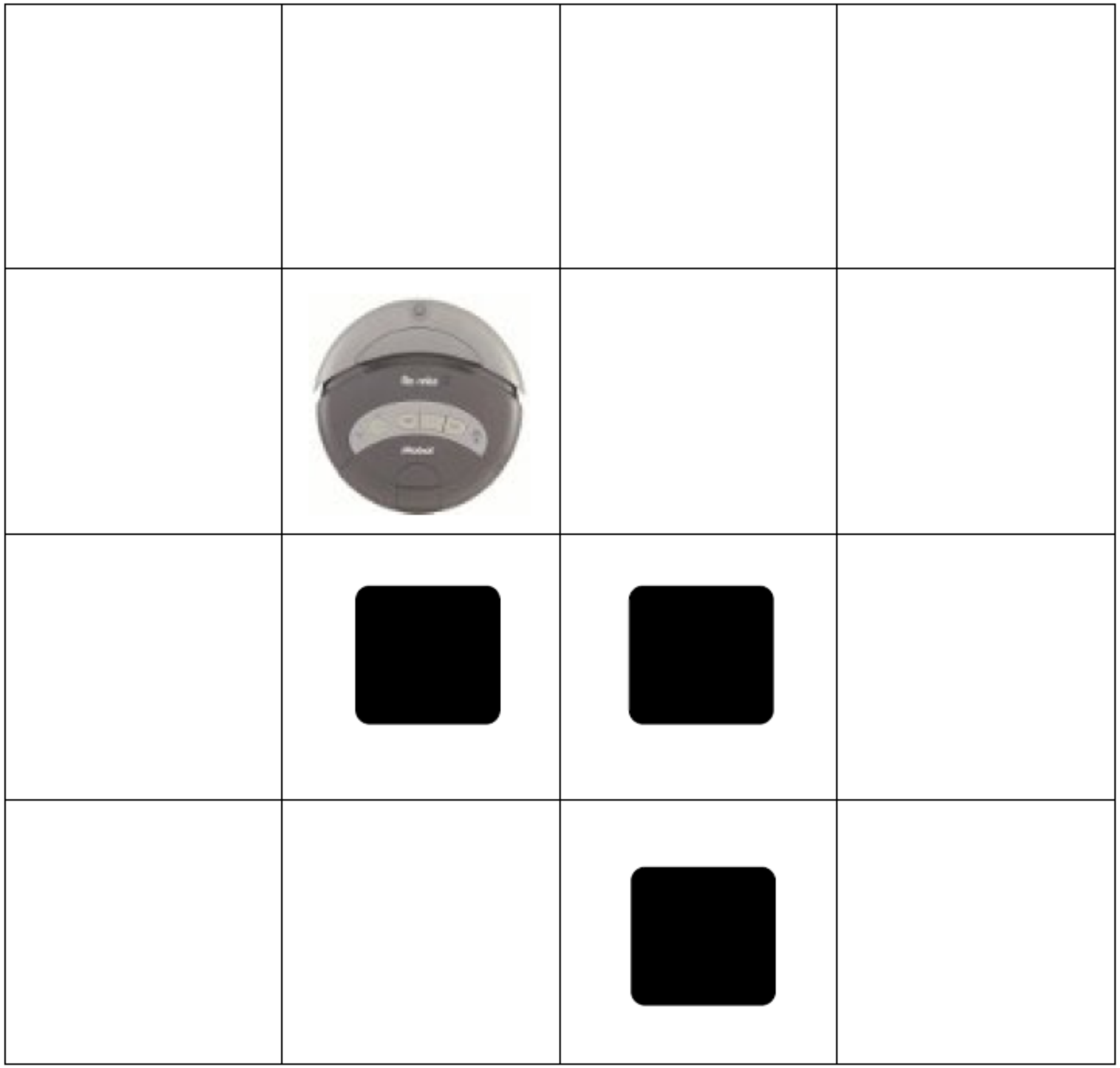
A run of an agent in an environment



A run of an agent in an environment

East!

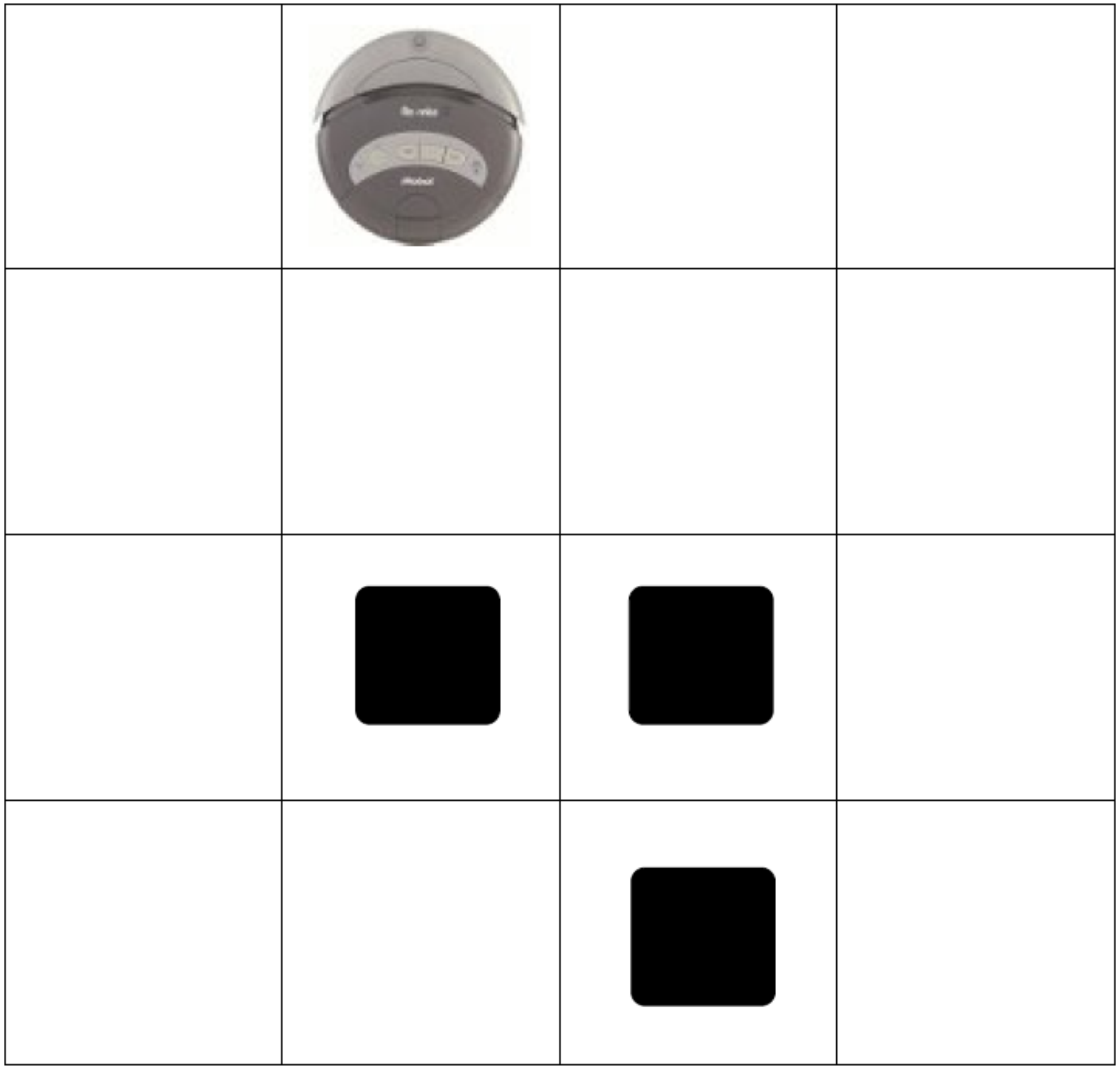
A run of an agent in an environment



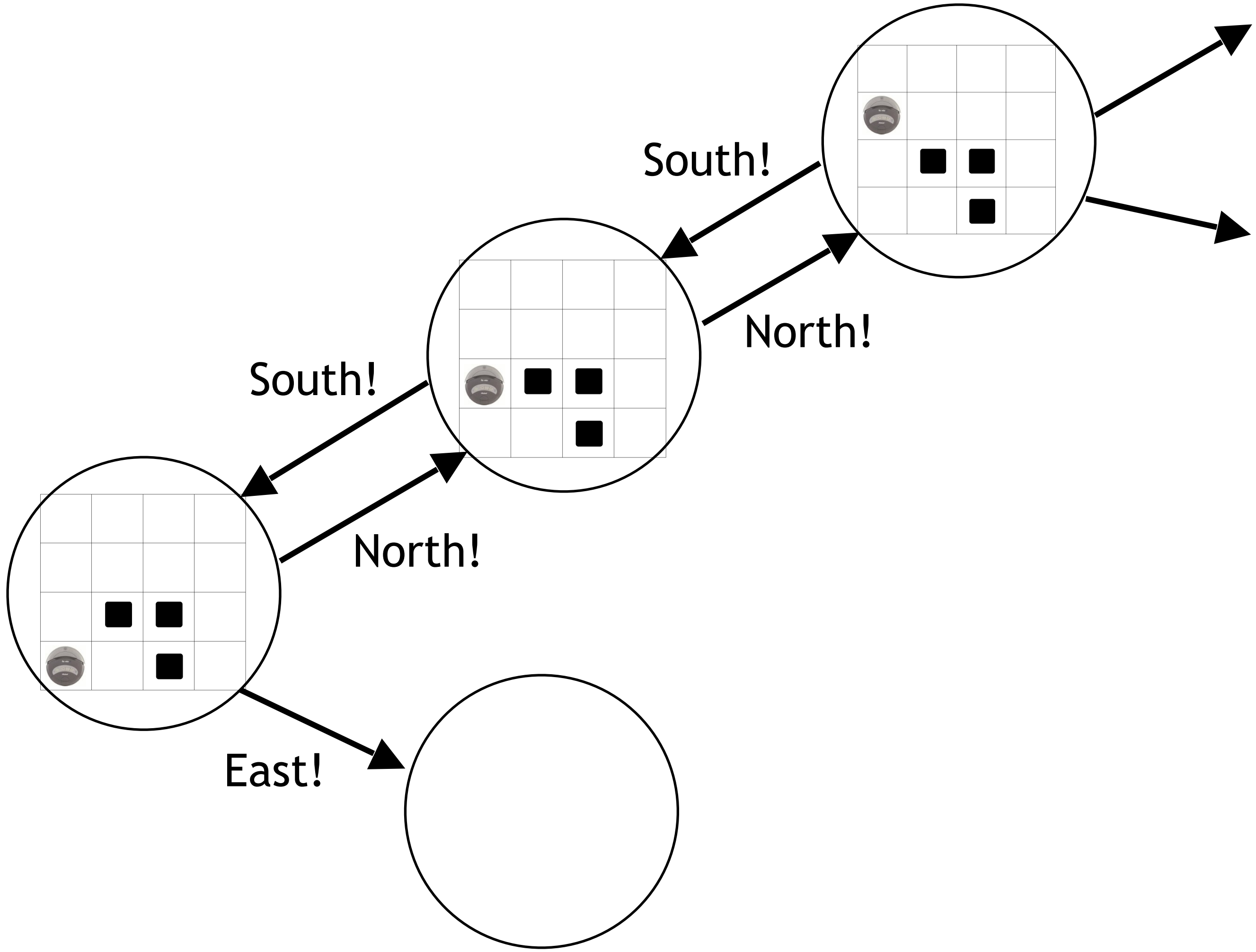
A run of an agent in an environment

North!

A run of an agent in an environment

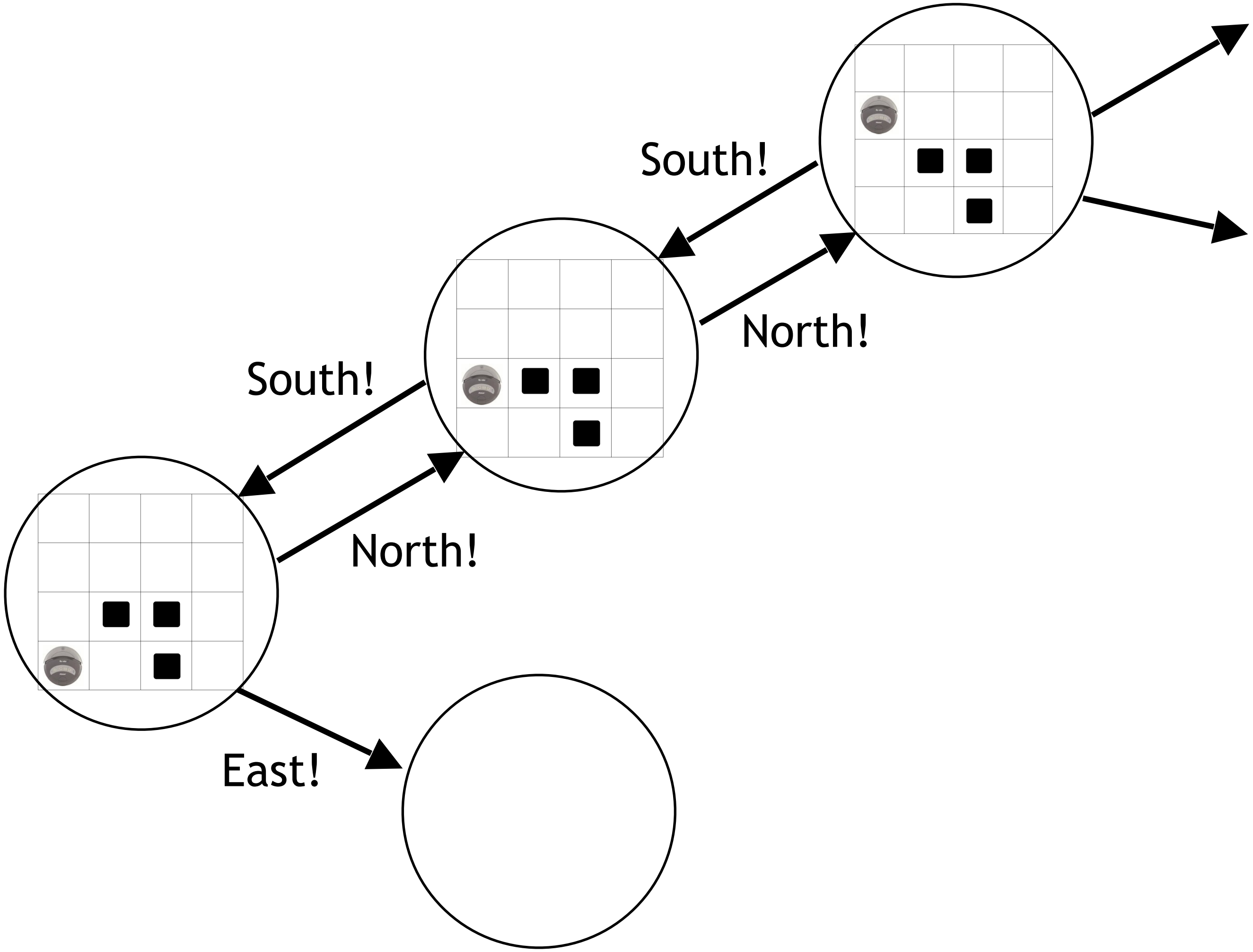


All possible runs of an agent from a start state



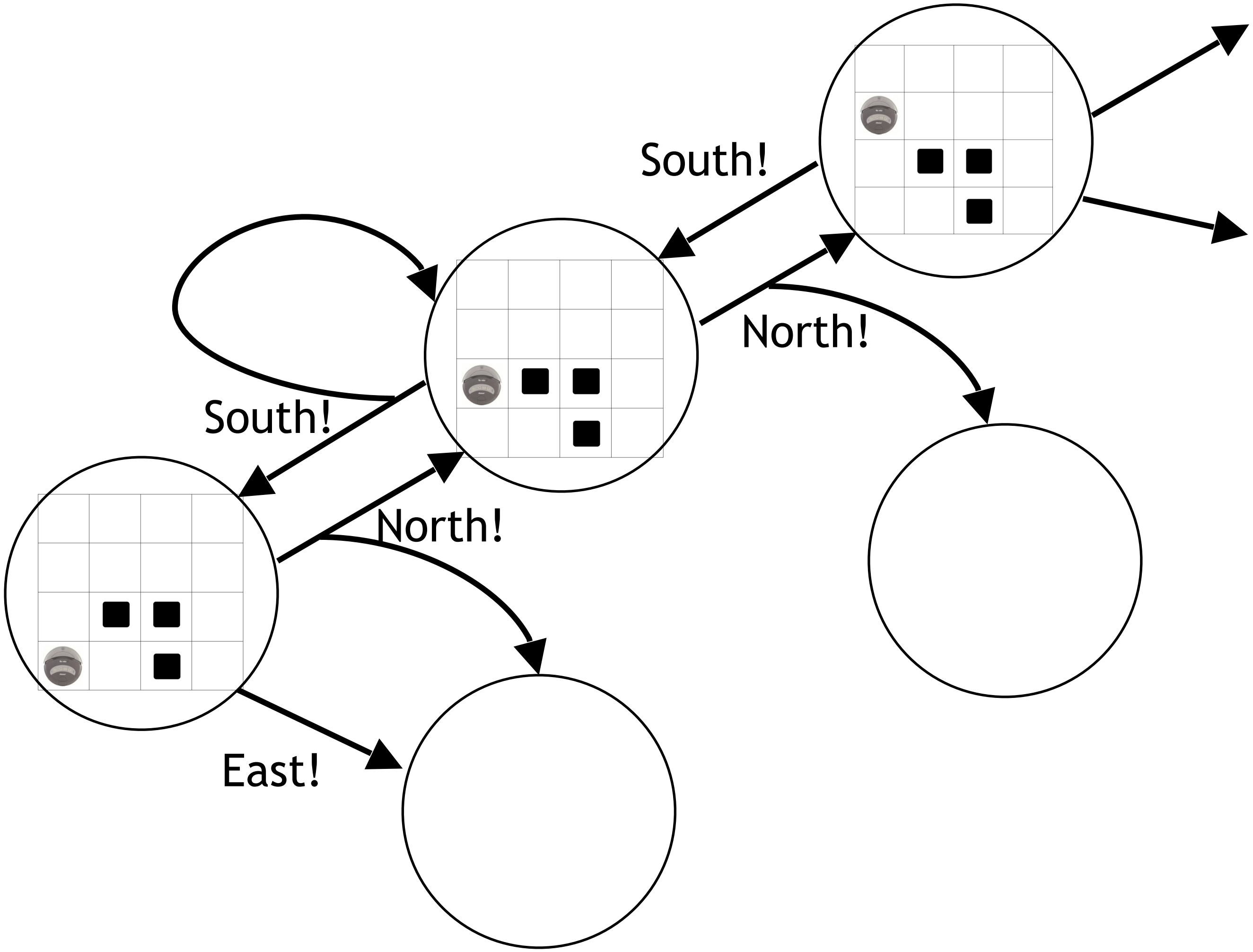
We can visualise the space of all possible runs.

All possible runs of an agent from a start state



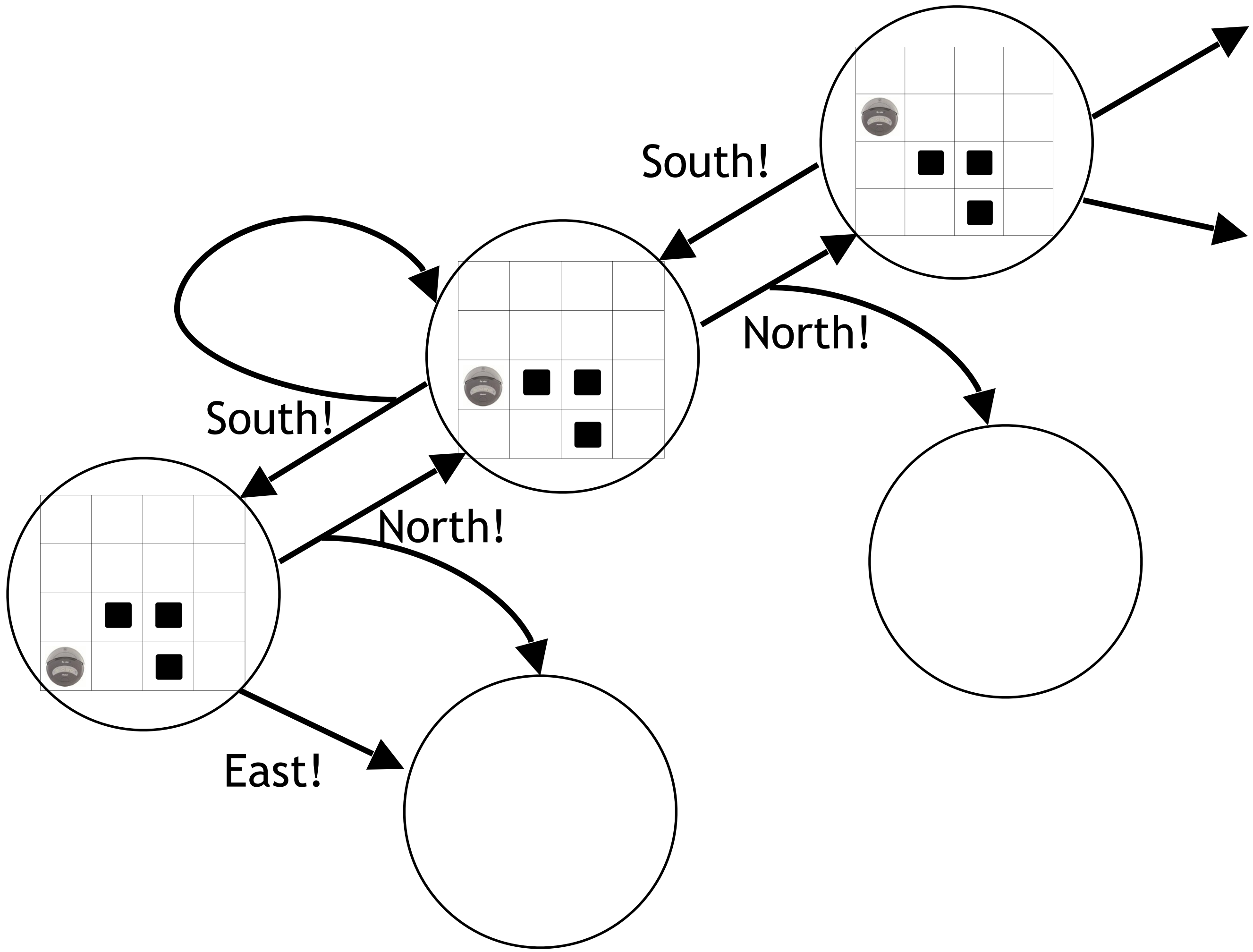
<u>Class activity</u>: What assumptions are we making here about the environment?

Runs in more complex environments



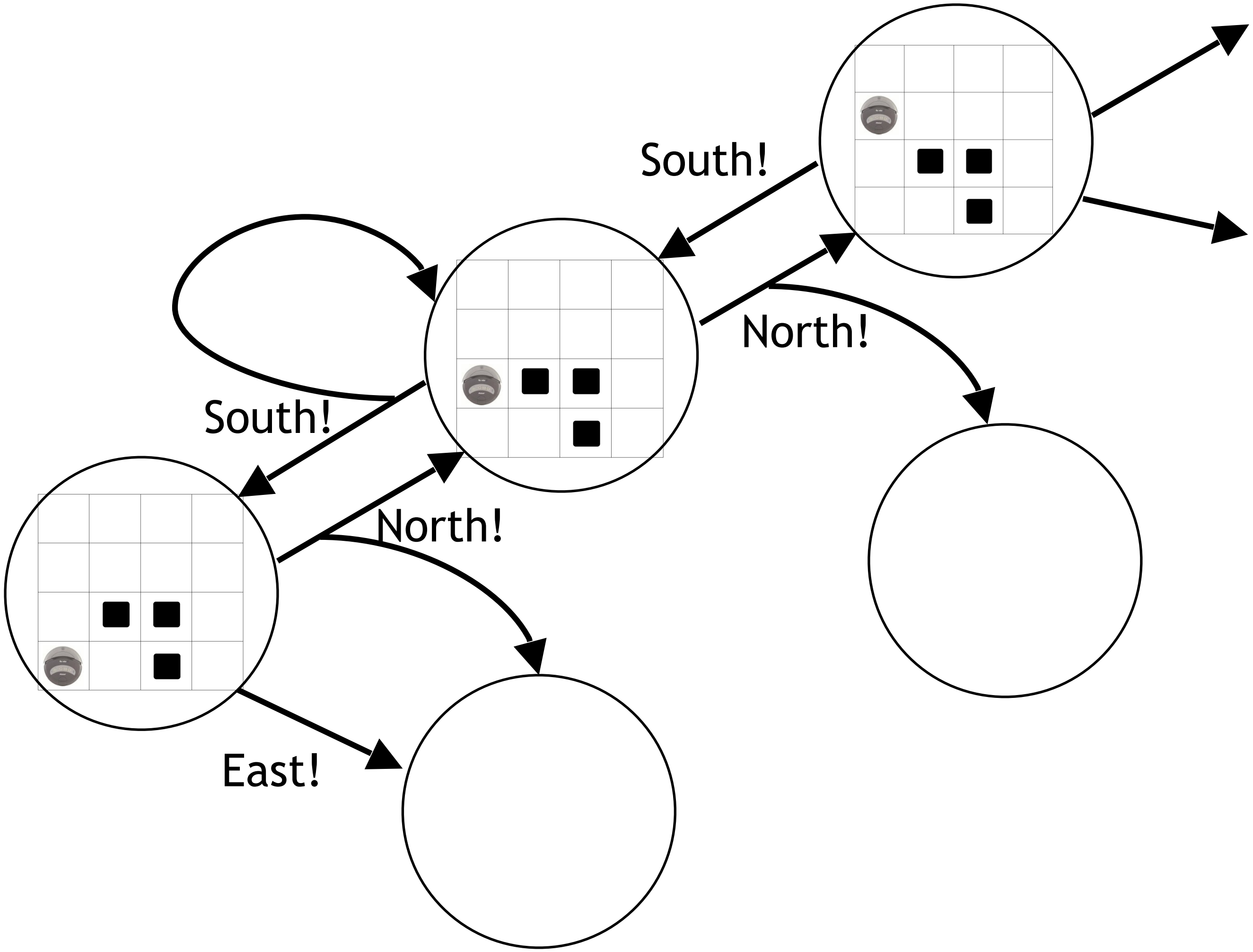
If environment is non-deterministic or dynamic then set of possible runs is more complex.

Runs in more complex environments



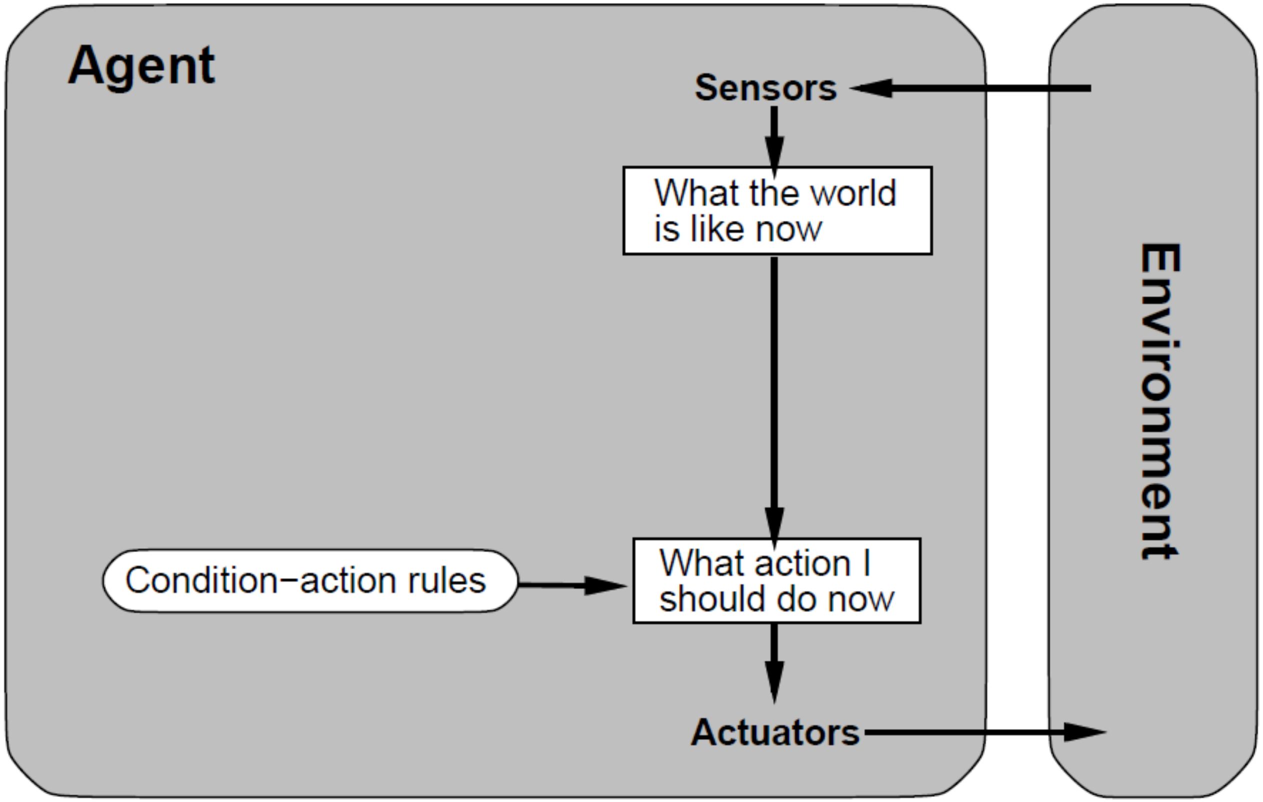
If environment is inaccessible there may be many possible start states to consider….

Runs in more complex environments



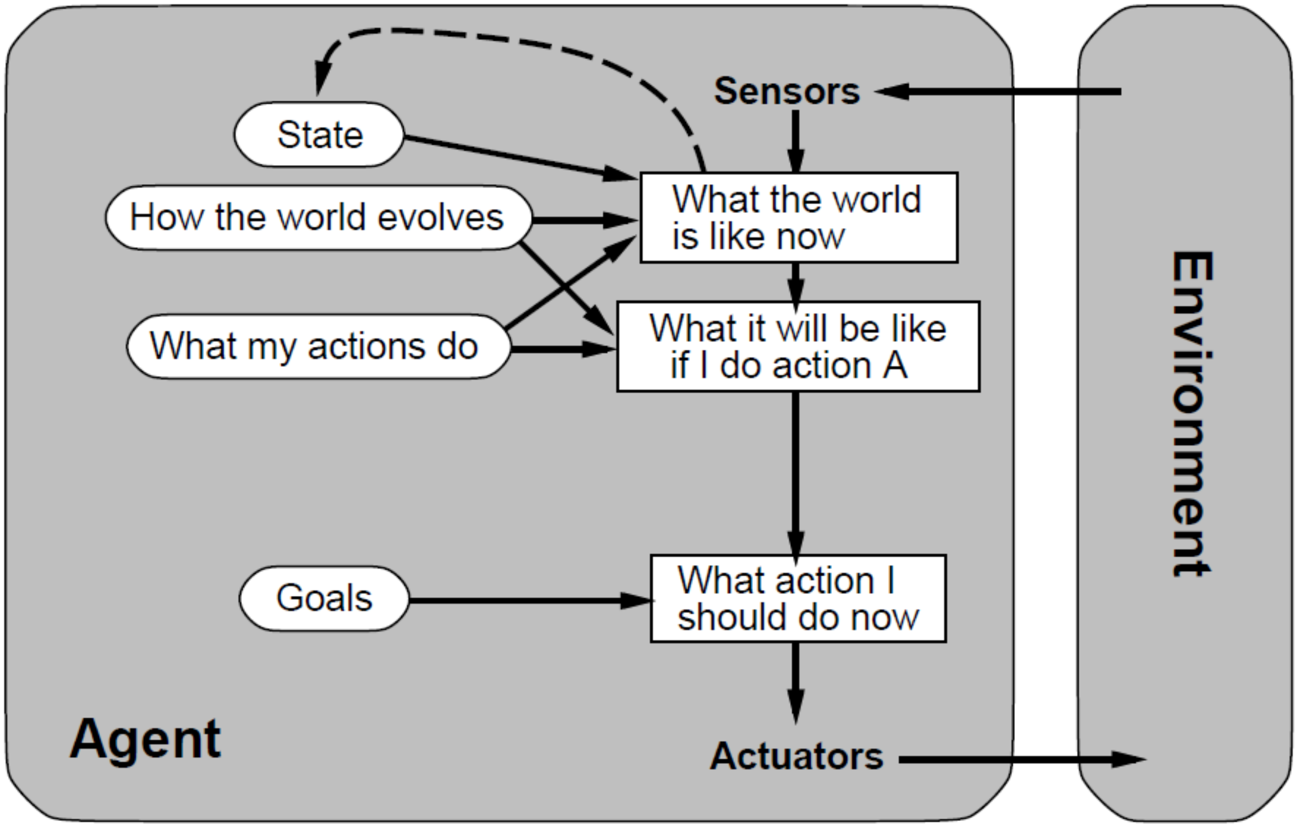
…and we may not be certain about which possible state we're in at any point.

When simple agents aren't enough



A simple agent maps percepts directly to actions.
Thus, it has to be able to make correct decision based only on the current percept.

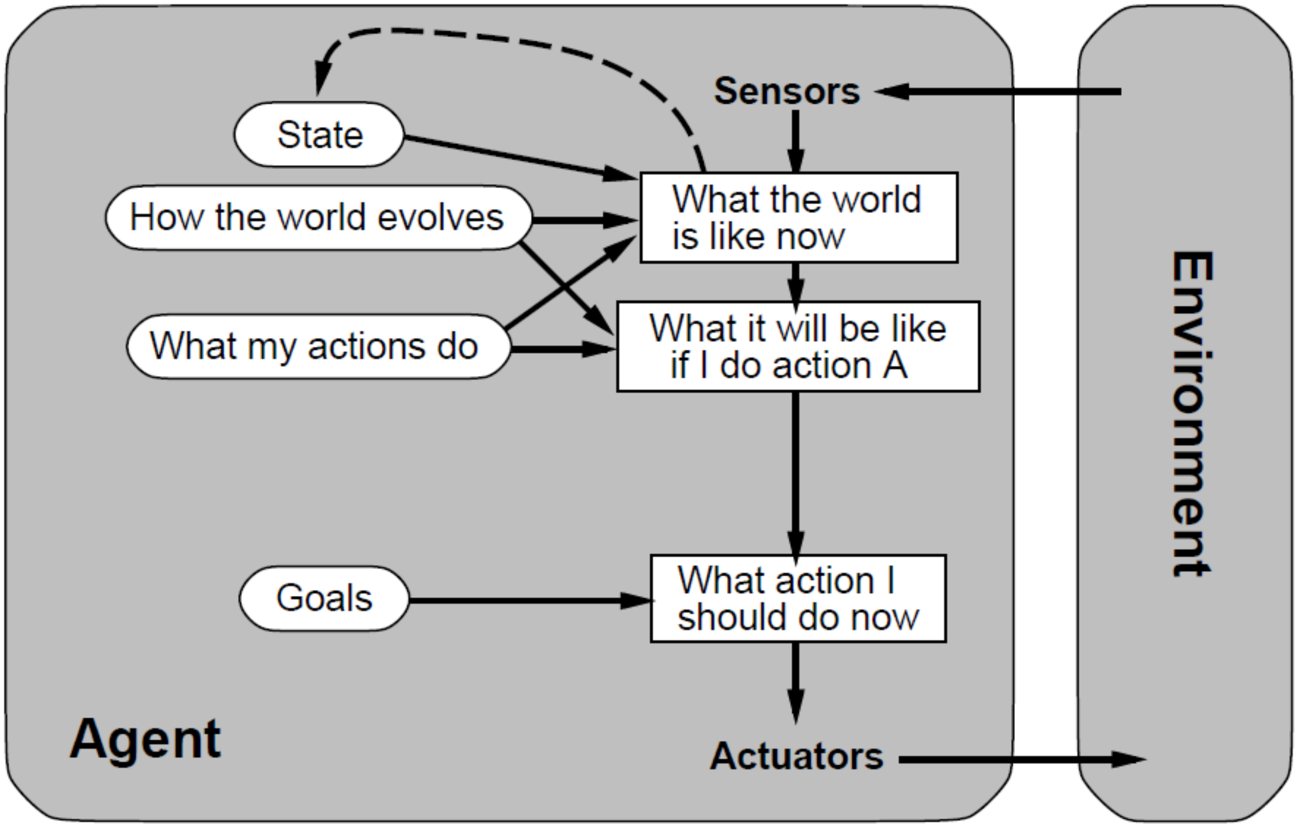
Agents with state and goals



New percept interpreted in light of existing knowledge about the state.

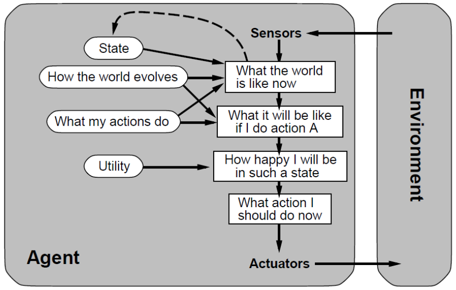
Knowledge about the world used to model inaccessible bits of the environment.

Agents with state and goals



Knowledge about world used to predict state resulting from action A.
Goal information used to identify desirable states.

Agents with state and goals

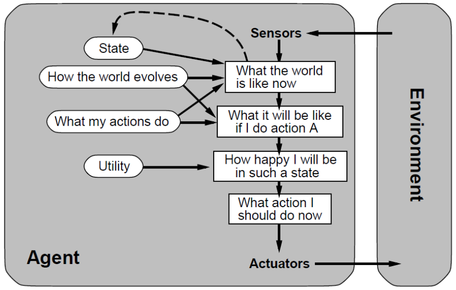


Agent has a well defined set of goals.

Is is easy for the agent to decide what to do?

No, goals may conflict. There be lots of different ways to achieve same goal.
Different likelihoods of successfully achieving goals.

Utility-based agents



Utility is a numerical value, tells us how “good” a state s is.
We denote this by $utility(s)$.

What if we don’t know for sure in which state we’ll end up in?

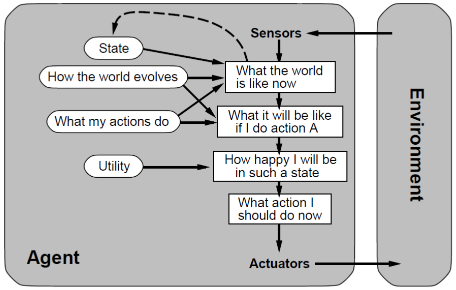
The expected utility of a particular state s is the probability that we’ll end up at s times by the utility of that state, i.e., $E[U_s] = utility(s) \cdot \Pr(State = s)$

The expected utility is the sum the sum of the expected utilities of all the possible resulting states, i.e.,

$$E[U] = \sum_{s \in States} E[U_s] = \sum_{s \in States} utility(s) \cdot Pr(State = s)$$

How to determine utility of a state not necessarily straight-forward!
Difficult to take a long term view.

Utility-based agents



A possible way of defining an optimal agent is such that it maximises $E[U]$. This can be done by changing the probability distribution $Pr(\cdot)$ (we will see examples later).

This is not always the best way to define an optimal agent!
Depending on the problem, e.g., maximising the minimal possible utility can be better…

How to determine utility of a state not necessarily straight-forward!
Difficult to take a long term view.

How might we determine a good outcome?

Credit: Xavier Caré

Stay tuned for the next lecture!

Summary

- We consider AI from the perspective of how an intelligent agent decides what to do.
- This lecture we looked at:
  - The notion of intelligent agents.
  - A classification of agent environments.
  - Some of the challenges for designing agents that can decide the right thing to do.
  - Some of the ethical questions around AI.
- We'll look at how some (advanced) AI techniques can be used to decide the right thing to do.