INTRODUCTION TO THE COURSE

dictionary.com meaning of ‘autonomous’ and ‘agent’

- **agent**:
  1. one that *acts* or has the power or authority to act.
  2. one empowered *to act* for or represent another.
  3. a means by which something is *done or caused*.
  4. a force or substance that *causes a change*.
  5. a representative or official of a government or administrative department of a government.
  6. a spy.
  7. (in linguistics) the noun noun phrase that specifies the person through whom or the means by which an *action effected*.

- **autonomous**:

  1. *not controlled* by others or by outside forces.
  2. independent in mind or judgement or government; *self directed.*
  3. *

     3.1. independent of the laws of another state or government; *self-governing.*

     3.2. of or relating to a self-governing entity.

     3.3. self-governing with respect to local or internal affairs.

Agents and autonomy: in our context

- **Agent**: An *entity* that acts, or does/perform/execute actions and causes/effects change.
  - entity: a software, hardware or an embedded system.
- **Autonomy**
  - Degree of autonomy to be bound by the constraint that 'it may not make us its slave'.
  - I.e., we have some control over it.
    * We program a behaviour and the agent stays within that.
    * We allow it to choose and perform actions but within a limit and it stays within that.
    * It can take high-level directives and goals and it can figure out what actions it needs to take and execute those actions.
  - But it is not micro-managed by us.

Expectations from autonomous agents

- Can understand high-level directives.
- Knows what actions it can do and what changes this action would cause.
- Able to sense and record its observations.
- Has some knowledge about the environment.
- Has the reasoning ability to figure out what it needs to do to achieve its goal.
- Has the ability to explain its observations to abduce previously unknown information.
- Has the ability to verify the correctness of its control.
- Has the ability to figure out its current situation.
- Has the ability to realize that its original plan of action is no longer valid and has the ability to make new plans from its *current situation.*
- Has the ability to diagnose, make diagnostic and repair plans.

- Our agents are mostly knowledge driven, but can learn some aspects from their interaction with the environment. (Disclaimer: We mostly overlook many other learning aspects of an agent.)

High level structure of the course

- Representing and reasoning about actions, in terms of their effects on the environment.
- Specifying and comprehending goals and directives.
- Verification of complex plans and agent architectures.

REPRESENTING AND REASONING ABOUT ACTIONS AND THEIR IMPACT

Representation Issues

- How to represent the transition diagram/graph/function? How to do the various kinds of reasoning (planning, plan verifications, observation assimilation, etc.)?
- For n fluents there are 2^n possible states.
- Even when actions are deterministic, a direct transition would require a table of $(2^{\#\text{of fluents}} \times \#\text{ of actions })$ entries.
- Alternative approach(es):
  - (i) List (what we know of) the effect of actions on the environment and any other relevant knowledge.
  - (ii) Assume 'Inertia': Nothing changes unless there is a reason to.
  - Approach 1: Explicitly compute valid states and the transition between states due to actions based on (i) and (ii).
  - Approach 2: Adapt the reasoning based on the transition graph to reasoning about a knowledge base consisting of (i) and (ii); avoid considering fluents that do not play a role in the reasoning task at hand.

- ∗ **But still need to define the transition function to validate the correctness of the reasoning.** – Semantics

- The more elaborate (i) is, the more difficult it is for computing the transition function, but the easier it becomes for the user to express (i), and the smaller the representation becomes.

The language $\mathcal{A}$: motivation

- Goal: A simple language to describe the effect of actions on the environment.
- Example:
  - *loading the gun* causes the gun to be *loaded.*
  - *shooting* causes the turkey *to die* if the gun is *loaded.*

    Or

    shooting causes the turkey *to die*

    and

    shooting can be executed only if the gun is *loaded.*
  - actions: load, shoot.
  - fluents (binary variables that describe the environment): loaded, alive
  - fluent literals: fluents and their negations (loaded, $\neg$ loaded, alive, $\neg$ alive).

Syntax of $\mathcal{A}$: an example

- load **causes** loaded.
- shoot **causes** $\neg$ alive **if** loaded.

  OR

  shoot **causes** $\neg$ alive.

  executable shoot **if** loaded.
- General form to represent effect of actions:
  - effect proposition: a **causes** f **if** p1, ... pn.
  - executability condition: **executable** a **if** q1, ... qm.

    a is an action; and f, p1, ..., pn, q1, ... qm are fluent literals.
- Domain description (D): a set of effect propositions, and executability conditions.

Syntax of $\mathcal{A}$ – cont.

- What else do we need besides describing effect of actions on the environment?
- What (kind of reasoning) will an autonomous agent do with this knowledge?
- Observations!
  - The turkey is initially alive.
  - The turkey was observed to be dead after shooting.
- Useful in planning, knowledge assimilation.
  - Planning from a complete initial situation: The gun is initially not loaded and the turkey is initially alive. How to achieve the state where the turkey is not alive?
  - Conformant planning with possibly incomplete initial state: We don't know if the gun is initially loaded or not and the turkey is initially alive. How to achieve the state where the turkey is not alive?
  - Knowledge assimilation: We don't know if the gun is initially loaded or not and the turkey is initially alive. We observe that the turkey is dead after shooting. What can we conclude more?

Syntax of $\mathcal{A}$: Observation and Queries?

- Observations (O):
  - f **after** a1, ..., an.
    - ∗ f was observed to be true after a1, ..., an is executed in the initial situation.
    - ∗ We are told that f would be true if a1, ..., an were to be executed in the initial situation.
  - **initially** f.
- Queries (Q): f **after** a1, ..., an.
  - Will f be true after executing a1, ..., an in the initial situation.
  - Would f be true if a1, ..., an were executed in the initial situation.
- Main reasoning relation (D, O) $\models$ Q.

  Defining the above is our next goal. (Semantics!)

Semantics of $\mathcal{A}$: motivating examples.

- D1 = { load **causes** loaded; shoot **causes** $\neg$ alive **if** loaded }
- D2 = { load **causes** loaded; shoot **causes** $\neg$ alive;
  executable shoot **if** loaded }
- O1 = { **initially** $\neg$ loaded; **initially** alive }
- O2 = { **initially** alive }
- O3 = = { **initially** alive; $\neg$ alive **after** load, shoot }
- Plan verification (hypothetical reasoning): (D1, O1) $\models$ $\neg$ alive **after** load, shoot?
- Simple planning: (D1, O1) $\models$ $\neg$ alive **after** α?
- Conformant planning: (D1, O2) $\models$ $\neg$ alive **after** α?
- Observation assimilation: (D1, O3) $\models$ **initially** loaded?
- Exercise: Replace D1 by D2 in the above reasoning tasks.