Autonomy in MAS: a classification attempt

Cosmin Carabelea

Ecole Nationale Supérieure des Mines de Saint-Etienne
Universitatea “Politehnica” Bucuresti

carabelea@cs.pub.ro

Plan

- Introduction
  - why autonomy?
  - terminology
  - the Vowels approach
- Related work on autonomy
- What is an autonomous agent?
- Autonomy in agents' architectures
- Conclusions

Introduction – why autonomy?

- "An agent is a real or virtual entity … that exhibits an autonomous behaviour." [Demazeau]
- "An intelligent agent is a computer system capable of flexible and autonomous action in some environment." [Wooldridge]

The autonomy is a defining characteristic of an agent, but there isn't a commonly agreed definition for it!

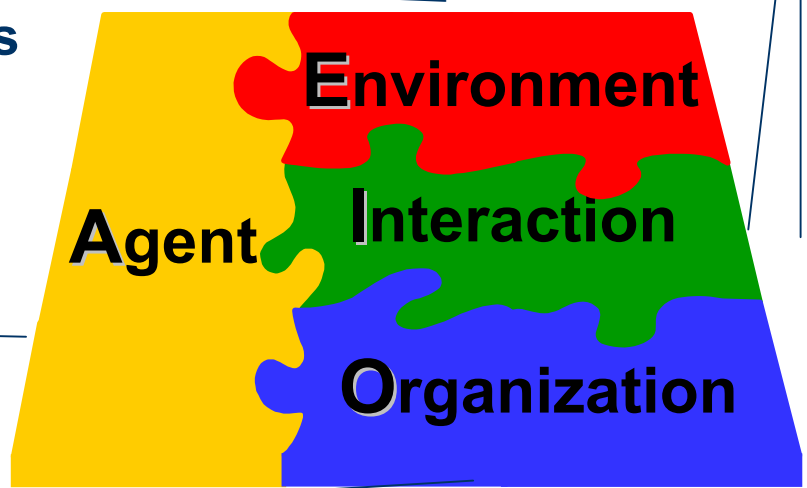
Introduction – terminology

- We are interested in goal-directed agents: their behaviours are guided by internal representations of the effects.
- A goal is a state of the world as represented in an agent's mind, which the agent wants to become true.
- A motivation is a higher-level notion that also guides the behaviour of an agent; it doesn't describe a state of the world but it is used to generate goals.
- An agent can delegate goals to or adopt goals from other agents.

The Vowels approach [Demazeau]

Common environment for the agents (Signals, resources, static/dynamic, …)

Acting entities (internal architectures, planning, …)



Interactions between agents (ACL, interaction protocols, …)

Relations between agents (organizational structures, norms, …)

Plan

- Introduction
- Related work on autonomy
- What is an autonomous agent?
- Autonomy in agents' architectures
- Conclusions

Related work on autonomy (1)

- The agent's autonomy from the user for making a decision. Usually the agent is a personal assistant.
- *Tambe et al.:* the agents use MDP to learn when to pass the control to the user and when not.
- The continuous passing of control to and from the user is called adjustable autonomy.
- Should the user allow the agent to make important decisions? Should the agent be able to refuse the user?

Related work on autonomy (2)

- A frequent usage of autonomy is with the meaning of social autonomy (an agent is autonomous from other agents).

- *Luck, d'Inverno*: an autonomous agent decides to adopt or not the goal from another agent based on its own motivations.
- *Barber*: autonomous agents vote for a common goal to pursue: an agent can vary from non-autonomous (doesn't vote) to master (it decides by itself).

Related work on autonomy (3)

- *Castelfranchi*: social autonomy is often (but not always!) related to the delegation or the adoption of goals: an autonomous agent is able to refuse a goal delegation from another agent.
- *Castelfranchi* used the dependence theory as a base for an attempt to unify different views on social autonomy.

- *Hexmoor* considers autonomy should always be studied in its situation (context) and not in a general, theoretical, manner.

Related work on autonomy (4)

- The use of autonomy introduces a degree of non-determinism in the behaviour of a multi-agent system.
- *Lopez y Lopez et al.:* norms have been proposed as a mean to restrain the autonomy of the agents using punishments and rewards.
- Problems with norm representation: the norm's context, the norm's addressee, the normative agent, etc. Moreover, there isn't a unique perspective on norms.

Related work on autonomy (5)

- Problems arise if there are norm-autonomous agents (*Dignum et al., Verhagen et al.*), agents able to decide if to obey or not a norm.
- *Dignum et al., Castelfranchi et al.,* have proposed agent architectures that take into account the norms in the system.
- *Verhagen* also proposed a classification of the autonomy in two categories and identified several levels of autonomy.

Plan

- Introduction
- Related work on autonomy
- What is an autonomous agent?
  - properties of autonomy
  - {U,I,O,E,A}-autonomy
  - a comprehensive definition
- Autonomy in agents' architectures
- Conclusions

What is an autonomous agent?

- With so many perspectives on autonomy, is it possible to find a comprehensive definition?
- We don't know yet, but there are some commonly agreed properties of autonomy that we will present next.
- We will then attempt to classify different forms of autonomy using the Vowels approach to which we add another dimension: the User.

Properties of autonomy

The object of autonomy (e.g. to make or not a decision, to adopt or not a goal, to obey or not a norm, etc.)

- The relative aspect of autonomy:

X is (not) autonomous from Y for p in the context C.

The agent

The influencer of the autonomy (e.g.: the user, another agent, the norms)

The agent can be autonomous in a situation and not autonomous in another

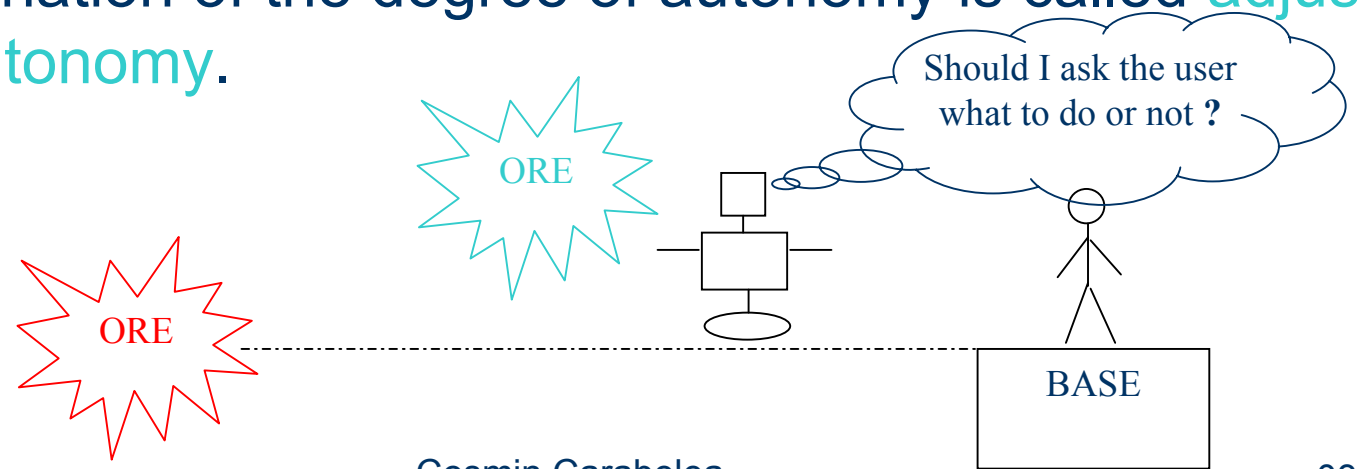
The agent is or is not autonomous. What makes it so?

- External vs. internal view of autonomy.

How does the agent adapt its behaviour using its autonomy?

U-autonomy

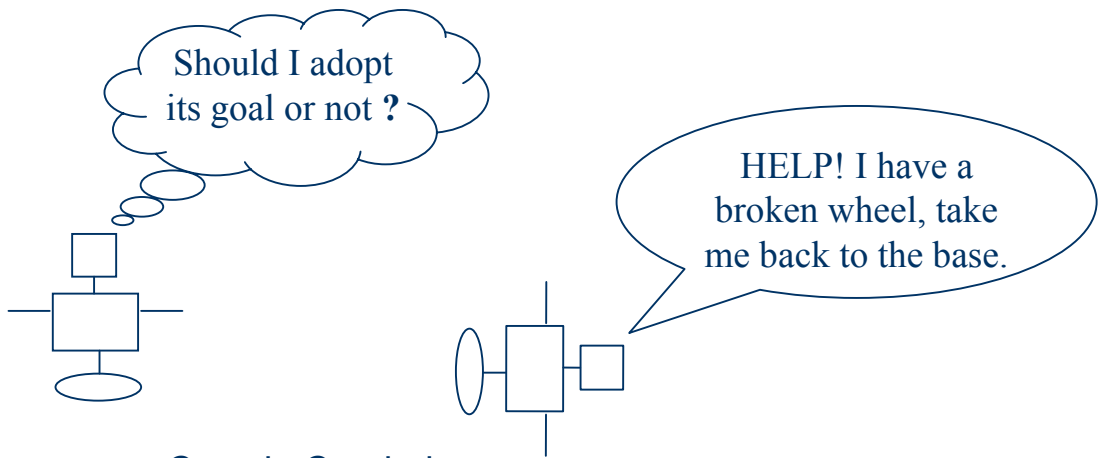
- *An agent is autonomous from* the user *for* making a decision *if it can* ***decide*** without the user's intervention.
- An agent can vary from a completely user-independent one to a completely user-dependent one. The dynamic variation of the degree of autonomy is called adjustable autonomy.



I-autonomy (social autonomy)

- *An agent is autonomous from* another agent *for* the adoption of a goal *if it can* ***decide*** to refuse the adoption of the goal from the other agent.

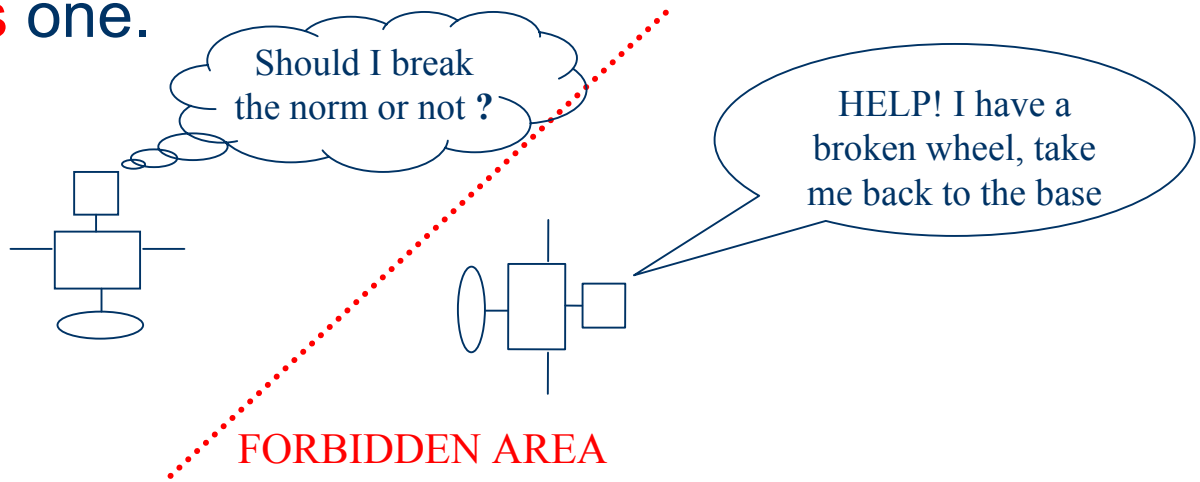
- An agent can vary from an autistic one to a benevolent one.



O-autonomy (norm autonomy)

- *An agent is autonomous from* a norm (*for* obeying it) *if it can* ***decide*** not to obey it.

- An agent can vary from a completely obeying one to a rebellious one.

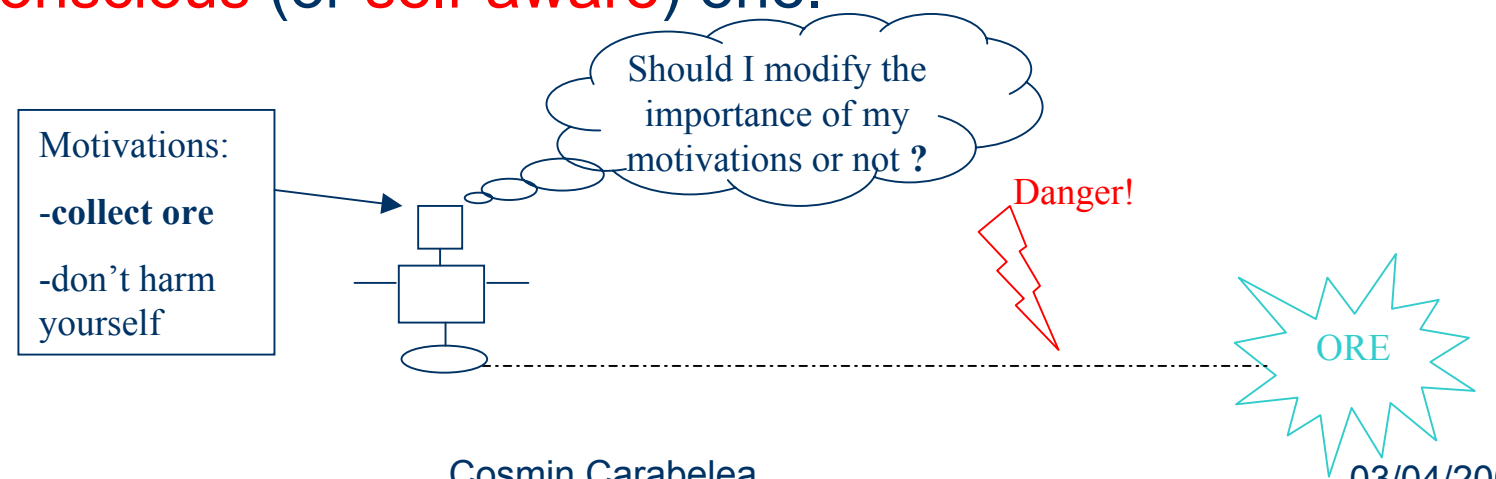


E-autonomy

- All agents are E-autonomous: they are situated in an environment, but not controlled by it.
- The “Descartes problem”: responses of many living systems to the environment are ‘neither caused by, nor independent of the external stimuli’
- Depending on how this response to the environment is formed, an agent can vary from a reactive to a deliberative one.

A-autonomy (self-autonomy)

- *An agent is autonomous from* itself *for* one of its motivations (emotions) *if it can* ***decide*** to modify (the importance of) that motivation (emotion).
- An agent can vary from an unconscious to a conscious (or self-aware) one.

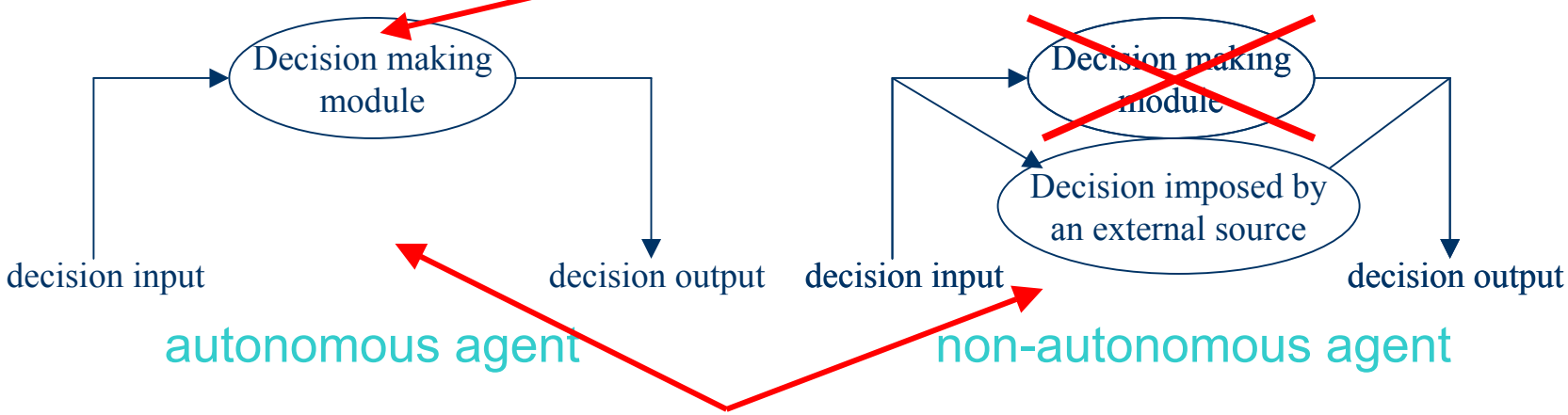


A comprehensive definition

An agent X is autonomous from Y for p in a context C if in C, X can make a local decision regarding p.

Local decision = independent of Y.

Internal perspective: *how* it will make that decision

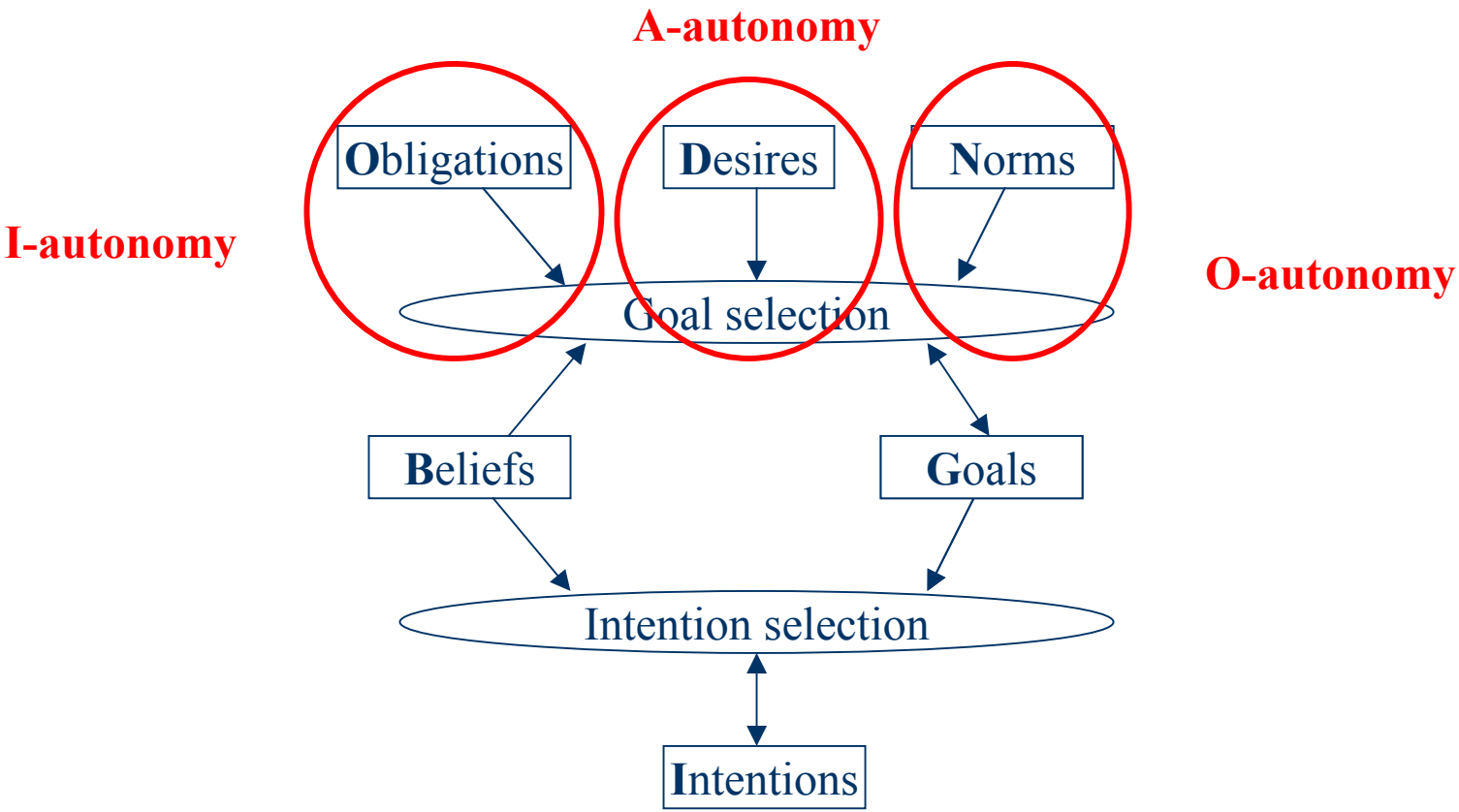


External perspective: it can or it cannot make the decision

Plan

- Introduction
- Related work on autonomy
- What is an autonomous agent?
- Autonomy in agent's architectures
  - B-DOING
- Conclusions

B-DOING [Dignum et al.]



Plan

- Introduction
- Related work on autonomy
- What is an autonomous agent?
- Autonomy in agent's architectures
- Conclusions

Conclusions

- Although it is a central notion in MAS, the autonomy doesn't have a commonly agreed definition.
- We have classified the different forms of autonomy using the Vowels approach and we have given a comprehensive definition of the autonomy.
- The need of an architecture for agents with adjustable autonomy (3A architecture) to identify what are the agent's parts that give it the autonomous character.