Knowledge Engineering

Semester 2, 2004-05

Michael Rovatsos

mrovatso@inf.ed.ac.uk

School of informatics

THE UNIVERSITY OF EDINBURGH

Lecture 10 – Agents and Multiagent Systems

15th February 2005

Where are we?

Last time ...

- Knowledge synthesis in closed systems

Today ...

- Basics of agents and multiagents systems
- Connection to knowledge engineering subject
- Discuss key features and research issues

Why agents?

- This is a knowledge engineering course
  What is the relationship to agent technology?
- So far concentrated on **expert systems** view characterised by
  - Capturing expert knowledge about a complex domain
  - Focus on inference and discovery of new knowledge
  - **Disembodiment** – no direct connection to action in an environment
- Agents as part of knowledge synthesis with an **open systems** outlook
  - How to bring knowledge from autonomous sources together
  - Different from merging (say) logical rules or knowledge bases

The Open Systems Perspective

- Current trends in computing (applications):
  - Ubiquity, interconnection, intelligence, delegation, human-orientation
- Examples:
  - Electronic markets, mobile/ubiquitous computing, peer-to-peer systems, grid computing, humanoid robots, Semantic Web etc.
- Computing environments become more and more **open**:
  - Interaction between heterogeneous software components
  - Components may enter or leave the system
  - Design in general not accessible from outside
  - Deployed by different persons/organisations with different goals
  - Certain kinds of behaviour cannot be enforced/precluded

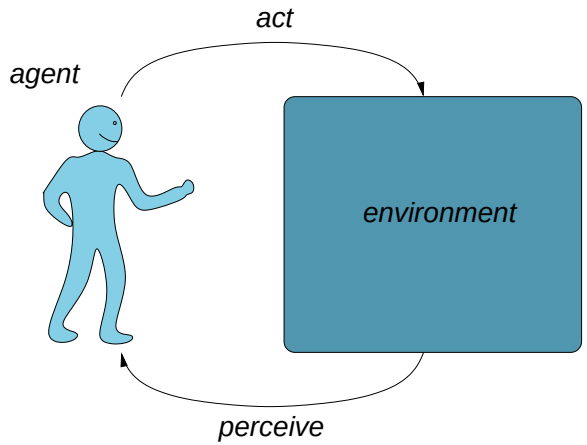
A Revolution in Computing?

- Interaction of automous components contradicts traditional engineering paradigm
  - If I don't own a component, how can I control it?
  - Requires new way of thinking from IT specialists
  - Correspondence to interaction in human society: communication rather than control
- Regulatory mechanisms must concentrate on interaction (rather than functionality of components)
  - Try to influence others' behaviours through own actions
- (Deeper) reasons for paradigm shift:
  - Global availability of IT resources
  - Use of IT beyond community of experts
  - Increasing complexity and anthropomorphism of software

What is an agent?

- Most widely accepted definition:

  An agent is anything that can perceive its environment (through its sensors) and act upon that environment (through its effectors)



- Autonomous agent:

  A computer system that is capable of independent (autonomous) action on behalf of its user or owner

Intelligent agents

- Intelligent agents are usually considered to be
  - autonomous (capable of independent action)
  - situated (embedded in an environment)
  - reactive (responsive to changes in environment)
  - proactive (able to take initiative for action)
  - rational (goal-oriented)
  - socially capable (able to communicate and interact with others)
- (Optional) additional features:
  - adaptiveness, mobility, lifelike qualities, real-time behaviour, sensorimotor capabilities, etc.

Agent Autonomy

Autonomous, situated in an environment, proactive and "intelligent" (in a way), but is it an agent?

Agent Autonomy

- Most disputed, but most essential aspect of agency
- Autonomy is a prerequisite for
  1. delegating complex tasks to agents
  2. ensuring flexible action in unpredictable environments
- Different definitions highlight different aspects
- A system is autonomous, ...
  - if it requires little help from the human user
  - to the extent that its behavior is determined by its own experience
  - if it can choose its own goal and the way to achieve it
  - if we have no direct control over it
  - if we don't understand its internal workings
- Autonomy is not a black and white notion!
- **Autonomy dilemma**: how to make the agent smart without losing control over it

Rationality: Reactivity vs. Proactiveness

- Example: The dung beetle

  After digging its nest and laying its eggs, it fetches a ball of dung from a nearby heap to plug the entrance; if the ball of dung is removed from its grasp en route, the beetle continues on and pantomimes plugging the nest with the nonexistent dung ball, never noticing that it is missing (quoted from Russell & Norvig)

- Truly flexible autonomous behaviour is hard to achieve!
- Trade-off between the two aspects because:
  - Environments are not fixed ➡ must be able to react to changes (involves monitoring own activity and environment, etc.)
  - Need for goal-oriented, planned activity ➡ not sufficient to respond to current circumstances

Rationality: Intentional Systems

- "The intentional stance": ascribing mental attitudes to machines
  - Beliefs, intentions, goals, free will etc.
- Do intentional models of computers make sense?
  - Does a thermostat have desires? Why not?
  - If we assume it does, will that buy us anything?
- Danger: Human tendency to "agentify" nature/artefacts
  - Praying to the rain god is quite similar to saying "This computer just doesn't like me"
- Arguments in favour of intentional models
  - Systems with human-like reasoning are easier to understand
  - For complex systems it might help reduce complexity
  - Usually not good idea if simpler model is available
  - Connection to autonomy: the less we know, the more well-suited is intentional stance

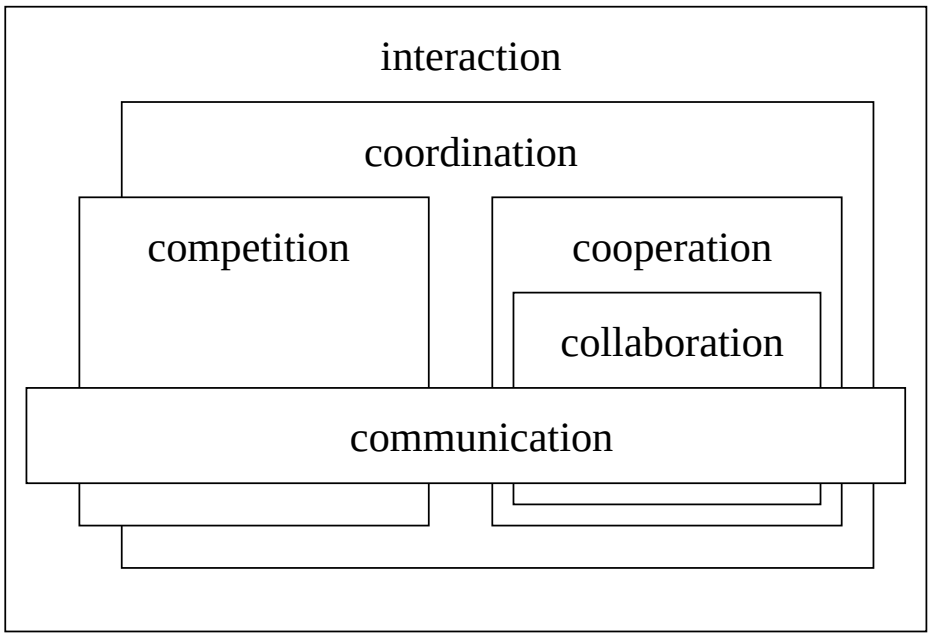
Social Ability

- In most real-world applications, environments are inhabited by multiple agents
- Each agent has limited resources/capabilities
  ➡ some goals may require others (not) to take action
- Social ability: the ability to manage one's interactions effectively
- Interaction and coordination:

  An interaction can be viewed as a formalisation of a concept of dependence between agents, no matter on whom or how they are dependent. Coordination is a special case of interaction in which agents are aware how they depend on other agents and attempt to adjust their actions appropriately.

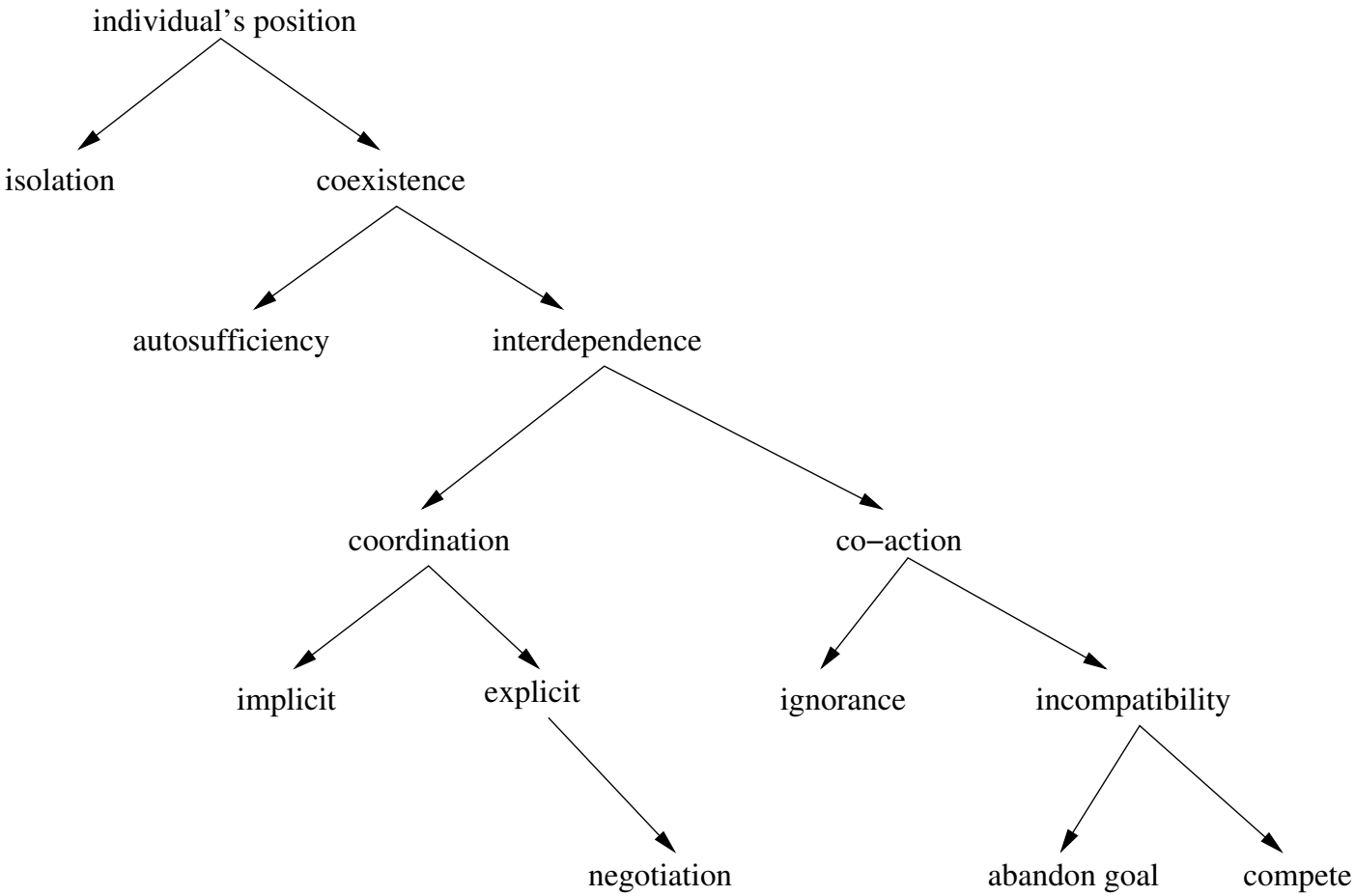
Social Ability

- Things to note:
  - Interaction does not always imply action
  - Coordination does not always imply communication
- Basic categorisation of interaction aspects:



Social Ability

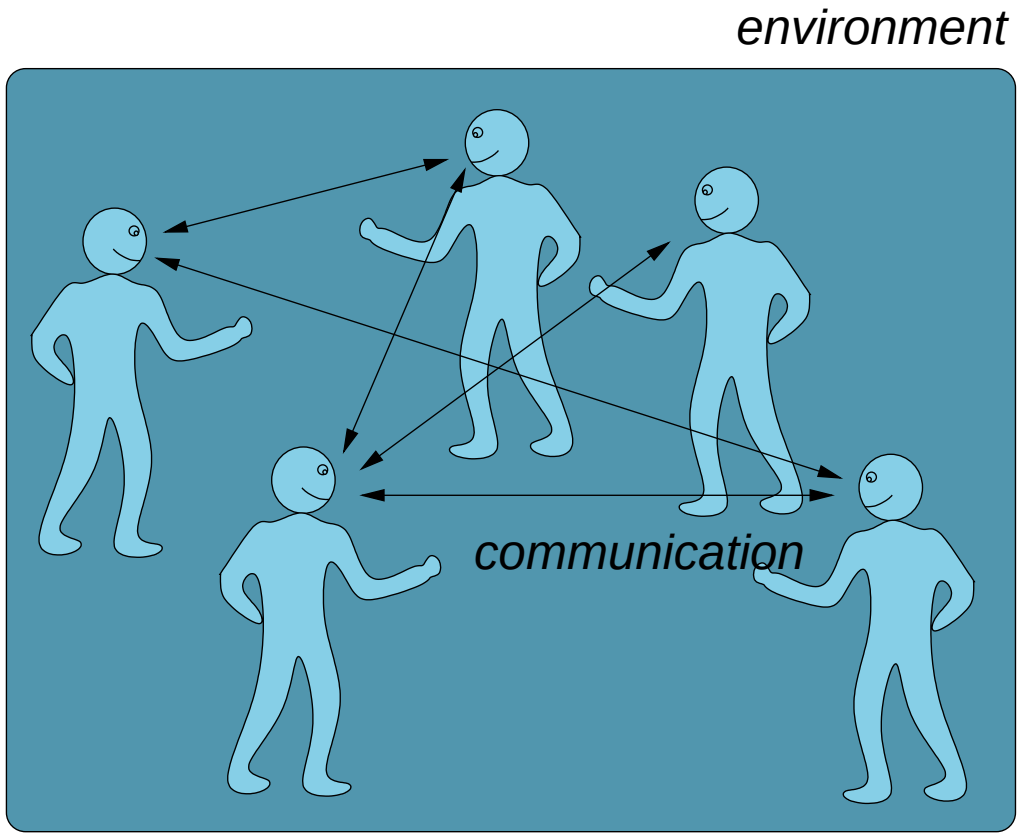
A more extensive interaction typology:



Multiagent Systems

- Collections of interacting agents situated in common environment
- Characterised by:
  - Computational distribution & decentralised control
  - Asynchronous action and communication
  - "Arm's length relationships" (no direct mutual manipulation)
  - Agents have incomplete information/limited capabilities
  - Mutual impact of agents' actions on goal achievement
- Two main categories of distributed AI (DAI) systems:
  1. Distributed problem-solving systems
     - strictly cooperative, benevolence assumption, centralised design, sometimes centralised control
  2. Multiagent systems
     - self-interested agents, heterogeneous/unknown agent design, decentralisation, openness

Multiagent Systems



Research Agenda

Essentially two key problems, most issues derived from this distinction btw **micro** and **macro** level:

1. **Agent design problem**
   - Agent architectures and reasoning mechanisms
   - Integration with sensing and action
   - Optimal agent design
2. **Society design problem**
   - Agent communication languages
   - Interaction protocols, conversation policies
   - Multiagent planning algorithms
   - Task and resource distribution mechanisms
   - Coordination mechanisms
   - Social norms and institutions

Some Sub-Areas

- Agent architectures
- Agent communication languages & protocols
- Negotiation & mechanism design
- Distributed (cooperative) problem solving
  - Search/planning/constraint satisfaction
- Multiagent learning & adaptation
- Agent-oriented software engineering
- Agent-based social simulation
- Research into different "types" of agents:
  - Robotic agents
  - Lifelike & believable agents
  - Interface agents
  - Information agents
  - Mobile Agents

The Programming Perspective

- Programming has evolved through different phases:
  - machine code ➡ assembly language ➡ machine-independent programming languages ➡ sub-routines ➡ procedures & functions ➡ abstract data types ➡ objects
- Are agents the next step?
  - "Objects will do it for free, agents will do it for money" (or because they "want" to)
- Intelligent agents and AI:
  - Buzzword used for lots of conventional software
  - Is building an intelligent agent not the ultimate goal of AI?
  - Maybe all we need is a system that will choose right action at right time (in a limited domain)

Summary

- Open systems perspective to knowledge synthesis
- Need for new methods ➡ agent abstraction
- Properties of agents: autonomy, rationality, social ability
- Philosophical issues: foundations of autonomy and intentionality
- Multiagent systems as spheres of agent interaction
- Different types of interaction
- Research agenda of agent technology
- Next time: **Agent Architectures**