

Intelligent & Autonomous Systems

Digital Culture for Digital Transformation

Andrea Omicini
andrea.omicini@unibo.it

DISI / BBS, Università di Bologna

Evening MBA Program, 16 April 2019



Parts

- I AI in Perspective
- II Beyond AI: Autonomy & Interaction



Outline of Part I. AI in Perspective

- 1 The Idea of AI
- 2 The Raise of AI
- 3 Main Themes in Early AI
- 4 The AI Winter
- 5 The Modern Era of AI



Outline of Part II. Beyond AI: Autonomy & Interaction

- 6 Autonomous Systems
- 7 Conversational Systems



Part I

AI in Perspective



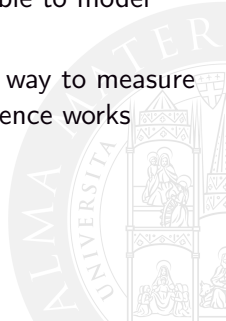
Next in Line...

- 1 The Idea of AI
- 2 The Raise of AI
- 3 Main Themes in Early AI
- 4 The AI Winter
- 5 The Modern Era of AI



The Idea of Intelligence

- understanding how intelligence works is a persistent issue for humans
- Aristotle's *logics* is the most outstanding example of that
- “understanding”, for humans, typically means to be able to model and reproduce
- building machines that can reproduce intelligence is a way to measure how much we do understand the way in which intelligence works



The Idea of Intelligent Machines

- humans as *makers* of machines exhibiting *intelligent behaviour*
- machines vs. humans—from The Turk [Standage, 2002] to Alpha Go [Silver et al., 2016]
- the basic question: *what is intelligence*, which we aim at capture and reproduce?



The Turing Test I

- after providing the fundamental theoretical machinery for computation (the Turing Machine [Turing, 1937]), Turing started exploring the limits of automatic computation: could it be a path towards intelligence of machines?
- the Turing Test [Turing, 1950]

"You are interacting via a computer keyboard and screen with something that is either another person or a computer program. The interaction is in the form of text – questions and answers. Your task is to determine whether the thing being interrogated is in fact a person. Now suppose, after some time, you cannot tell whether the thing is a person or program. Then, Turing argued, you should accept that the thing being interrogated has human-like intelligence." [Wooldridge, 2018]

The Turing Test II

- *behavioural* interpretation of intelligence—independent from how it works, dependent on what it can achieve
- indistinguishable from human behaviour
- it has limitations, it has been refined, it is anyway meaningful



Against AI I

- many scientists have argued against the chance of machine intelligence
- mostly philosophers, like Searle [Searle, 1980]

The Chinese Room

"Imagine a room in which a man, who understands no Chinese, receives, through a slot in the door, questions written in Chinese. When he receives a question, the man carefully follows detailed instructions written in English to generate a response to the question, which he passes back out through the slot. Now suppose the questions and responses are part of a Chinese Turing test, and the test is passed." [Wooldridge, 2018]

Against AI II

- interesting argument
- yet, so many assumptions and mistakes in it
- which does not mean its conclusions are wrong, of course



Next in Line...

- 1 The Idea of AI
- 2 The Raise of AI**
- 3 Main Themes in Early AI
- 4 The AI Winter
- 5 The Modern Era of AI



AI is Born in 1956

The Golden Age of AI

- Dartmouth College, New Hampshire, Summer School, 1956
- John McCarthy invites all scholars interested in *computing towards intelligence*
- among those
 - Marvin Minsky, co-founder of AI Lab at MIT
 - Alan Newell, Herb Simon, authors of Logic Theorist (an automatic theorem prover), likely the first AI program [Newell and Simon, 1956]
 - John McCarthy, inventor of LISP, the first programming language for AI [McCarthy, 1981]
- the term “Artificial Intelligence” was actually coined there, to describe the overall new field of research
- the dualism between AI as *intelligent behaviour* and AI as *intelligent process* already there, from the beginning

General AI I

General purpose AI

- building general-purpose intelligence machines is the goal of **General AI**
- we do have a *poor understanding* of human intelligence, and of intelligence in general
- early AI focussed then on *intelligent components*



General AI II

Components of intelligence

- perception
- machine learning
- problem solving & planning
- reasoning
- natural language understanding



General AI III

Perception

- understanding the **environment**
- through **sensors** of any sort
- *interpreting* the overall situation
- ! one of the most difficult task of AI

Machine learning

- **learning** from **data**
- building models (e.g., classification)
- making *predictions*
- e.g., face recognition through training

General AI IV

Problem solving & planning

- devising a course of actions towards a *goal*
- based on a *repertoire of actions*
- e.g., playing games

Reasoning

- representing knowledge
- *inferring new knowledge* from available one
- in a *consistent* and *robust* way



General AI V

Natural language understanding

- ability to understand human languages
- either spoken or written
- possibly engage in conversations with humans



Next in Line...

- 1 The Idea of AI
- 2 The Raise of AI
- 3 Main Themes in Early AI**
- 4 The AI Winter
- 5 The Modern Era of AI



Search

Searching for solution

- from the current state of the world to a desired one, called **goal**
- based on a repertoire of available *actions*
- modelled in terms of their
 - *pre-conditions*, or, the state of the world in which they can be applied
 - *effects*, or, the state of the world they bring about
- a *search tree* can be built
 - from the *initial state* as its root—*forward search*
 - from the *final state* as its root—*backward search*



The Problem of Search

Explosion of the space of solutions

- search trees become too large
- *combinatory explosion*
- e.g., game of chess
- possible approach: *heuristics* as rule of thumb to direct search and reduce computational complexity
- e.g., IBM's Deep Blue [Campbell et al., 2002] against Gary Kasparov in 1996, combining computational power with heuristics



Planning in the Block World

SHRDLU [Ward, 2006]

- the Blocks World is a *virtual environment* where a number of coloured objects (blocks, boxes, pyramids) are arranged on a surface (a table)
- SHRDLU was aimed at arranging the objects following user's instructions
- using a simulated robot arm with heuristics
- it worked: both the *user interaction* via a sort of *natural language*, and the *practical reasoning* (reasoning towards action)

limits *language* and *environment* actually very *simplistic*, not representative of *real-world interaction*



Robots

SHAKY with STRIPS [Furcy and Koenig, 2006]

- a remotely-controlled mobile robot
- moving objects in an office-like environment by pushing them
- planning with STRIPS
- it somehow worked
- limits: limited ability to perceive and understand real-world environment, so the environment had to be built around it to make SHAKY work
- basically, it *looked* like a real-world robot, but *never worked* in the real world



Next in Line...

- 1 The Idea of AI
- 2 The Raise of AI
- 3 Main Themes in Early AI
- 4 The AI Winter**
- 5 The Modern Era of AI



Complexity I

Microworld does not scale up to real-world

- solutions in microworlds apparently do not work in real world
- qualitative or quantitative issues?

NP-complete problems

- *Non-deterministic Polynomial time complete* (NP-complete) problems
 - easy to check correctness of solutions
 - almost impossible to *find* a solution in a reasonable amount of time
 - e.g., Travelling Salesman Problem (TSP) [Papadimitriou, 1977]
- ! almost all relevant problems in AI are (at least) NP-complete



Complexity II

AI Winter

- AI started with way too much hype
- e.g., General Problem Solver (GPS) never worked for real-world problems [Newell, 1959]
- in a couple of decades, nothing was really accomplished
- ! Funding for AI research were cut everywhere, starting from the seventies



Tackling Complexity with Knowledge I

A way out

- humans do not explore enormous search spaces to find solutions
- they use (possibly extensive) **domain knowledge** to *cut down the search space*
- e.g., chess players
- ! the idea: exploiting explicit (domain) knowledge representation to make solutions computable



Tackling Complexity with Knowledge II

Expert systems

- a new class of AI systems emerged
 - no longer General AI, since they were tailored to *specific domains*
 - working only there
 - yet finally facing real-world problems successfully
- *expert systems* were built by
 - *eliciting knowledge* from domain experts
 - representing them in terms of *rules*
 - able to provide *answers to human questions* over a domain
 - along with an *explanation*
 - often coping with *uncertainty*



Tackling Complexity with Knowledge III

MYCIN [Shortliffe, 1976]

- nearly 200 rules
- about blood disease
- working as a doctor's assistant
- issues
 - difficult to extract knowledge from humans, and to precisely formalise it in terms of rules
 - operating only in a very narrow area of expertise



Tackling Complexity with Knowledge IV

CYC [Lenat et al., 1985]

- representing everything in the world in terms of rules
- an expert systems containing all the common sense human knowledge in terms of rules
- Doug Lenat's CYC started in 1984...
- ... and is still alive
- way too ambitious: some call it a failure
- yet it now lives out of its commercial applications



Computational Logic for AI

Logic Programming

- capturing knowledge, inferring new knowledge—with the precision and rigour of logic
- computing with logic: the PROLOG programming language [Colmerauer and Roussel, 1996]—nowadays usually written as Prolog
 - knowledge and programs expressed in terms of *logic axioms* (Horn clauses)
 - altogether forming a *logic theory*
 - against which *goals* are demonstrated
 - *computation as goal demonstration* [Lloyd, 1984]
- issues
 - heavy computational load
 - not fitting (alone) real-world representation and computation

Bayesian reasoning

Coping with uncertainty

- humans are not good decision makers
- *rational decisions* better than human decisions in most fields, in particular where uncertain information is available
- e.g., when *probabilistic reasoning* is involved
- Bayes developed the math in the eighteen century, and a couple of centuries were required to make *Bayesian reasoning* usable for AI
[Barber, 2012]
- e.g., many NLP systems use Bayesian networks



Nouvelle AI

Against reasoning, against representation [Brooks, 1991a, Brooks, 1991b]

- frustration from absence of real-world results led to focus on real world itself
- against knowledge-based and logic-based AI
- **perceive and react** as the main cycle of robots in the real world, rather than *(symbolic) reasoning*



Behavioural AI

Robotics revolution [Brooks, 1986]

- focus on the expected specific intelligent behaviours
- just build them, according to how they do relate with each other
- in particular, which ones take the precedence on the others
- e.g., Roomba vacuum cleaner



Next in Line...

- 1 The Idea of AI
- 2 The Raise of AI
- 3 Main Themes in Early AI
- 4 The AI Winter
- 5 The Modern Era of AI



Autonomous Vehicles

Grand DARPA Challenges

- race for autonomous vehicles in the desert of Nevada (2005)
- won by STANLEY [Thrun et al., 2006], a converted Volkswagen Touareg, equipped with seven onboard computers, interpreting sensor data from GPS, laser rangefinders, radar, and video feed



Machine Learning I

Computer that learn

- machine learning (ML) aims at computers that learn knowledge they have not be explicitly told [Michalski et al., 1983]
- they are typically *trained*



Machine Learning II

Supervised learning

- training by examples of the thing that the computer is trying to learn—*training data*
- e.g, face recognition

Reinforcement learning

- a system is able to experiment by *making decisions*
- receives *feedback* on those decisions
- and learns from that



Machine Learning III

Three factor for the success of ML

- *scientific breakthroughs*—deep learning dealing with complex problems
- training requires *lots of data*—nowadays data are hugely available
- training requires computational power—nowadays computational power is more and more available



Neural Nets

Nets of neurons

- *neural nets* implement ML by connecting many artificial neurons in complex networks [Lippmann, 1987]
- each *neuron* gets input from neighbours and produces an output based on some *weights*
- learning occurs by adjusting *weights*
- two-phase research
 - first, simple neural networks show potential and are proven to have limits
 - then, more complex networks were proven able to overcome these problems
- e.g., character recognition
- ! issue: opaque, no explanation made available—knowledge hidden in weights

Triumph of ML

Alpha Go [Silver et al., 2016]

- in 2014 DeepMind demonstrated a system learning how to play arcade games just looking at the video and accessing the scores, using the same controls as humans
- acquired by Google, they built Alpha Go, which beat Go champion Lee Sedol 4 to 1 in 2016
- exploiting *deep neural networks* along with *self-training*
- Go search space is so huge that brute force just does not work: so, it was considered impossible for a machine to beat a human
- so, this also made everybody know that there were no known limits to the ability that machine intelligent could reach

AI Today

AI everywhere [Wooldridge, 2018]

- digital assistants like Siri, Alexa, and Cortana are classic AI
- face recognition software in social media is AI
- the software in car's satellite navigation system is AI
- product recommendations from online stores are made by AI
- cruise control systems in modern cars are AI
- AI-based automated translation systems are routinely used across the world
- ...

AI is just **everywhere** already



Part II

Beyond AI: Autonomy & Interaction



Next in Line...

6 Autonomous Systems

7 Conversational Systems



Focus on. . .

6 Autonomous Systems

- **Premises**
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- Challenges



Before Autonomous Systems: Machines I

Constructing for understanding

- building machines with
 - initiative
 - autonomy
 - knowledge

for *understanding ourselves*, and the *world* where we live

- “playing God” to understand the world



Before Autonomous Systems: Machines II

Relieving humans from fatigue

- goal: **substituting human** work in
 - quality
 - quantity
 - cost
- more, better, cheaper work done
 - new activities become feasible
- which work?
 - first, **physical**
 - then, **repetitive**, enduring
 - subsequently, **intellectual**, too
 - finally, simply more **complex** for any reason—or, all reasons together

Before Autonomous Systems: Machines III

Some steps beyond

- delegating human functions to machines
 - within already existing social structures, organisations, and processes
- creating new functions
 - then, making new social structures, organisations, and processes possible
 - example: steam engines on wheels
- essentially, *changing the world we live in*

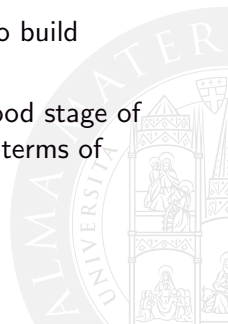


Autonomous Systems Around



Social Pressure

- activities that might be *delegated* to *artificial systems* grow in *number* and *complexity*
- people & organisations are already experiencing / conceiving “somehow autonomous” systems, and ask for more
- engineers are not yet trained on *general approaches* to build *autonomous systems*
- however, the theory of autonomous systems is at a good stage of development, and technologies are rapidly growing in terms of *availability* and *reliability*



Focus on. . .

6 Autonomous Systems

- Premises
- **What is Autonomy?**
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- Challenges



About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - **Autonomy in Language**
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Oxford Dictionary of English (2nd Edition revised 2005)

Etimology

Early 17th cent.: from Greek *autonomia*, from *autonomos* 'having its own laws', from *autos* 'self' + *nomos* 'law'.

Dictionary

autonomy

- the right or condition of self-government
- a self-governing country or region
- freedom from external control or influence; independence.
- (in Kantian moral philosophy) the capacity of an agent to act in accordance with objective morality rather than under the influence of desires

Oxford Thesaurus of English (2nd Edition revised 2008)

Thesaurus

autonomy

- self-government, independence, self-rule, home rule, sovereignty, self-determination, freedom, autarchy;
- self-sufficiency, individualism.



Merriam-Webster I

Dictionary

autonomy

- 1 the quality or state of being self-governing; especially: the right of self-government
- 2 self-directing freedom and especially moral independence
- 3 a self-governing state

synonyms accord, free will, choice, self-determination, volition, will

antonyms dependence (also dependance), heteronomy, subjection, unfreedom



Merriam-Webster II

Thesaurus

autonomy

- 1 the act or power of making one's own choices or decisions:
accord, free will, choice, self-determination, volition, will
- 2 the state of being free from the control or power of another:
freedom, independence, independency, liberty,
self-determination, self-governance, self-government,
sovereignty



About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - **Autonomy in Biology**
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Autonomy and Biological Systems

Living systems. . .

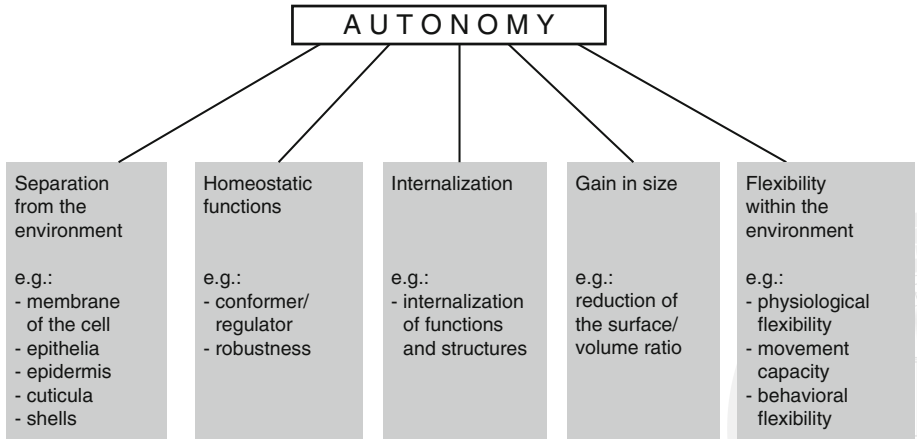
... are the first systems provided of (some level of) autonomy that we have knowledge of

- they *work* as **autonomous systems**
- they **evolved** to *become autonomous*

The study of autonomy in biological systems

- the hierarchy of living systems provide examples of many **different** levels of autonomy—from lower to higher levels of autonomy
 - the evolutionary view over biological systems potentially sheds light over the **role** of autonomy in living systems
- the study of living system may help us understanding the many different **sorts** of autonomy, and their **role** in (artificial) systems in general

Resources to Change Autonomy



[Rosslenbroich, 2014]

A Definition for Autonomy in Living Systems

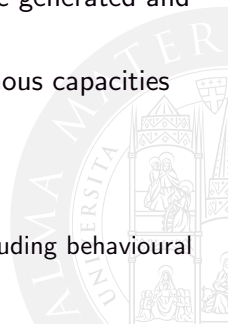
General autonomy [Rosslenbroich, 2014]

Living systems are *autonomous* in the sense that they *maintain themselves in form and function* within time and achieve a *self-determined flexibility*

- 1 they generate, maintain, and regulate an *inner network* of interdependent, energy-consuming processes, which in turn generate and maintain the system
- 2 they establish a *boundary* and actively regulate their *interaction* and exchange with the *environment*
- 3 they specify their *own rules of behaviour* and react to external stimuli in a *self-determined* way, according to their internal disposition and condition
- 4 they establish an *interdependence* between the system and its parts within the organism, which includes a *differentiation* in subsystems
- 5 they establish a *time autonomy*
- 6 they maintain a phenotypic stability (*robustness*) in the face of diverse perturbations arising from environmental changes, internal variability, and genetic variations

Increasing Autonomy in Evolution [Rosslenbroich, 2014]

- *increasing autonomy* is defined as an *evolutionary shift* in the *system-environment relationship*, such that
 - interactive autonomy* the direct influences of the environment on the respective individual systems are gradually reduced
 - constitutive autonomy* stability and flexibility of self-referential, intrinsic functions within the systems are generated and enhanced
- a set of resources can be involved to change autonomous capacities
 - ① changes in *spatial separation* from the environment
 - ② changes in *homeostatic* capacities and robustness
 - ③ *internalisation* of structures or functions
 - ④ increase in body *size*
 - ⑤ changes in the *flexibility* within the environment, including behavioural flexibility



Autonomy as a Driver for Evolution

- *autonomy is an essential trait of living systems*
- biological systems *evolve* towards *increasing* degrees of *autonomy*
- autonomy is an **evolutionary pattern**



About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - **Autonomy in Philosophy**
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Internet Encyclopedia of Philosophy I

Many acceptations of autonomy

- general** an individual's capacity for self-determination or self-governance
- folk** inchoate desire for freedom in some area of one's life
- personal** the capacity to decide for oneself and pursue a course of action in one's life
- moral** the capacity to deliberate and to give oneself the moral law, rather than merely heeding the injunctions of others
- political** the property of having one's decisions respected, honored, and heeded within a political context



Internet Encyclopedia of Philosophy II

Individual autonomy

- after Kant, autonomy is an essential trait of the **individual**, and strictly related with its **morality**, represented by some high-level ethical principles
- then, with the relation between its inner self and its individual actions
- that is, mind and behaviour



Internet Encyclopedia of Philosophy III

Independence from oneself

- a more demanding notion of autonomy requires not only self-determination, but also **independence from oneself**
- this conception is connected with notions of freedom and choice, and (maybe) non-determinism
- and requires the ability of reasoning on (and possibly changing) not just one own course of actions, but one own goals



About...

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - **Autonomy in Military Systems**
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Unmanned Systems Integrated Roadmap FY 2011-2036 I

Automatic vs. autonomous

automatic systems are fully pre-programmed and act repeatedly and independently of external influence or control. An automatic system can be described as self-steering or self-regulating and is able to follow an externally given path while compensating for small deviations caused by external disturbances. However, the automatic system is not able to define the path according to some given goal or to choose the goal dictating its path.

autonomous systems are self-directed toward a *goal* in that they do not require outside control, but rather are governed by laws and strategies that direct their behavior. Initially, these control algorithms are created and tested by teams of human operators and software developers. However, if *machine learning* is utilized, autonomous systems can develop modified strategies for themselves by which they select their behavior. An autonomous system is self-directed by choosing the behavior it follows to reach a human-directed goal.

Unmanned Systems Integrated Roadmap FY 2011-2036 II

Four levels of autonomy for unmanned systems [Edwards, 2013]

Various levels of autonomy in any system guide how much and how often humans need to interact or intervene with the autonomous system:

- human operated
- human delegated
- human supervised
- fully autonomous



Unmanned Systems Integrated Roadmap FY 2011-2036 III

Level	Name	Description
1	Human Operated	A human operator makes all decisions. The system has no autonomous control of its environment although it may have information-only responses to sensed data.
2	Human Delegated	The vehicle can perform many functions independently of human control when delegated to do so. This level encompasses automatic controls, engine controls, and other low-level automation that must be activated or deactivated by human input and must act in mutual exclusion of human operation.
3	Human Supervised	The system can perform a wide variety of activities when given top-level permissions or direction by a human. Both the human and the system can initiate behaviors based on sensed data, but the system can do so only if within the scope of its currently directed tasks.
4	Fully Autonomous	The system receives goals from humans and translates them into tasks to be performed without human interaction. A human could still enter the loop in an emergency or change the goals, although in practice there may be significant time delays before human intervention occurs.



Unmanned Systems Integrated Roadmap FY 2011-2036 IV

Autonomy & unpredictability

- the special feature of an autonomous system is its ability to be goal-directed in *unpredictable situations*.
- this ability is a significant improvement in capability compared to the capabilities of automatic systems.
- an autonomous system is able to *make a decision* based on a set of rules and/or limitations.
- it is able to determine *what information is important* in making a decision.
- it is capable of a *higher level of performance* compared to the performance of a system operating in a predetermined manner.



About...

6 Autonomous Systems

- Premises
- What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - **Autonomy in Cars**
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
- Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
- Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- Challenges



Autonomous Cars

So much hype

- “This is probably the biggest thing to hit the auto industry since the first car came off the assembly line”
- yet, it will probably take more time than hopeful ones expects, less time than skeptical one think

From automatic to autonomous, one step at a time

- most of the technologies will be implemented gradually
- leading to a slowly-yet-steadily *growing level of autonomy*

Levels of Autonomy for Cars I

An international standard [SAE J3016, 2018]

- it provides a taxonomy describing the full range of levels of **driving automation** in *on-road motor vehicles*
- it includes functional definitions for *advanced levels* of *driving automation* and related terms and definitions



Levels of Autonomy for Cars II

Levels

- 0 no automation — every aspect of driving is in the hands of the human driver
- 1 driver assistance — the car assists the human driver in some tasks—e.g., cruise control, lane-keeping
- 2 partial automation — *advance driver assistance system* (ADAS): some *automated functions work together* to transfer some control from the human driver (who remains in charge of the monitoring) to the car—e.g., Tesla Autopilot
- 3 conditional automation — *automated driving systems* (ADS): under some circumstances, the car can perform *all aspects* of the driving task—e.g., Audi AI traffic jam pilot
- 4 high automation — under certain conditions, the car can perform all driving functions, with no need for a human driver—e.g., Google self-driving car
- 5 full automation — under *any condition*, the car can perform *all driving tasks*; humans are just passengers

Levels of Autonomy for Cars III

Level	Name	Narrative definition	DDT		DDT fallback	ODD
			Sustained lateral and longitudinal vehicle motion control	OEDR		
Driver performs part or all of the DDT						
0	No Driving Automation	The performance by the <i>driver</i> of the entire <i>DDT</i> , even when enhanced by <i>active safety systems</i> .	<i>Driver</i>	<i>Driver</i>	<i>Driver</i>	n/a
1	Driver Assistance	The <i>sustained</i> and <i>ODD</i> -specific execution by a <i>driving automation system</i> of either the <i>lateral</i> or the <i>longitudinal vehicle motion control</i> subtask of the <i>DDT</i> (but not both simultaneously) with the expectation that the <i>driver</i> performs the remainder of the <i>DDT</i> .	<i>Driver and System</i>	<i>Driver</i>	<i>Driver</i>	Limited
2	Partial Driving Automation	The <i>sustained</i> and <i>ODD</i> -specific execution by a <i>driving automation system</i> of both the <i>lateral</i> and <i>longitudinal vehicle motion control</i> subtasks of the <i>DDT</i> with the expectation that the <i>driver</i> completes the <i>OEDR</i> subtask and <i>supervises</i> the <i>driving automation system</i> .	System	<i>Driver</i>	<i>Driver</i>	Limited

...

Levels of Autonomy for Cars IV

ADS ("System") performs the entire DDT (while engaged)			System	System	Fallback-ready user (becomes the driver during fallback)	Limited
3	Conditional Driving Automation	The <i>sustained</i> and <i>ODD</i> -specific performance by an <i>ADS</i> of the entire <i>DDT</i> with the expectation that the <i>DDT fallback-ready user</i> is <i>receptive</i> to <i>ADS</i> -issued requests to <i>intervene</i> , as well as to <i>DDT performance-relevant system failures</i> in other vehicle systems, and will respond appropriately.				
4	High Driving Automation	The <i>sustained</i> and <i>ODD</i> -specific performance by an <i>ADS</i> of the entire <i>DDT</i> and <i>DDT fallback</i> without any expectation that a <i>user</i> will respond to a <i>request to intervene</i> .				
5	Full Driving Automation	The <i>sustained</i> and unconditional (i.e., not <i>ODD</i> -specific) performance by an <i>ADS</i> of the entire <i>DDT</i> and <i>DDT fallback</i> without any expectation that a <i>user</i> will respond to a <i>request to intervene</i> .				Unlimited

Summary of levels of driving automation [SAE J3016, 2018]

ADS automated driving system

DDT dynamic driving task

ODD operational design domain

OEDR object and event detection and response

About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - **Autonomy in Social Sciences & AI**
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Autonomy as a Relational Concept [Castelfranchi, 1995] I

Autonomy as a social concept

- an agent is autonomous mostly in relation to other agents
- autonomy has no meaning for an agent in isolation

Autonomy from environment

- the Descartes' problem: (human, agent) behaviour is affected by the environment, but is not depending on the environment
- situatedness, reactivity, adaptiveness *do not* imply lack of autonomy



Autonomous Goals [Castelfranchi, 1995] |

Agency

- agents as teleonomic, **goal-oriented** entities
- that is, whose *behaviour* is *not casual* under any acceptance of the term



Autonomous Goals [Castelfranchi, 1995] II

Agents & goals [Conte and Castelfranchi, 1995]

- agents in a society can be generally conceived as either *goal-governed* or *goal-oriented* entities
 - goal-governed entities refer to the strong notion of agency, i.e. agents with some forms of cognitive capabilities, which make it possible to explicitly represent their goals, driving the selection of agent actions
 - goal-oriented entities refer to the weak notion of agency, i.e. agents whose behaviour is directly designed and programmed to achieve some goal, which is not explicitly represented
- in both cases, agent goals are *internal*



Autonomous Goals [Castelfranchi, 1995] III

Executive vs. motivational autonomy

executive autonomy — given a goal, the agent is autonomous in achieving it by itself

motivational autonomy the agent's goals are somehow self-generated, not externally imposed

Autonomy & autonomous goals

- autonomy requires *autonomous goals*
- executive autonomy is not enough for real autonomy

Goal-autonomous agent

A *goal-autonomous agent* is an agent endowed with its own *goals*

About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - **Autonomy in Programming Languages**
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Evolution of Programming Languages: The Picture

	Monolithic Programming	Modular Programming	Object-Oriented Programming	Agent Programming
Unit Behavior	Nonmodular	Modular	Modular	Modular
Unit State	External	External	Internal	Internal
Unit Invocation	External	External (CALLED)	External (message)	Internal (rules, goals)

[Odell, 2002]

Evolution of Programming Languages: Dimensions

Historical evolution

- monolithic programming
- modular programming
- object-oriented programming
- agent programming

Degree of modularity & encapsulation

- unit behaviour
- unit state
- unit invocation

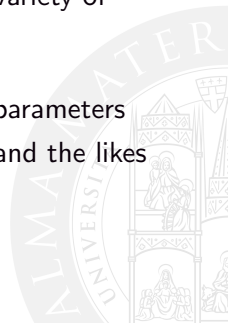
Monolithic Programming

- the basic unit of software is the whole program
- programmer has full control
- program's state is responsibility of the programmer
- program invocation determined by system's operator
- behaviour could not be invoked as a reusable unit under different circumstances
 - modularity does not apply to unit behaviour



Modular Programming

- the basic unit of software are structured loops / subroutines / *procedures* / ...
 - this is the era of procedures as the primary unit of decomposition
- small units of code could actually be *reused* under a variety of situations
 - modularity applies to subroutine's code
- program's *state* is *determined* by externally supplied parameters
- program invocation determined by CALL statements and the likes



Object-Oriented Programming

- the basic unit of software are *objects* & *classes*
- structured units of code could actually be *reused* under a variety of situations
- objects
 - have *local control* over variables manipulated by their own methods
 - variable state is *persistent* through subsequent invocations
 - object's state is *encapsulated*
 - are *passive*—methods are invoked by external entities
 - modularity does not apply to unit invocation
 - object's *control* is *not encapsulated*



Agent-Oriented Programming

- the basic unit of software are *agents*
 - encapsulating everything, in principle
 - by simply following the pattern of the evolution
 - whatever an agent is
 - we do not need to define them now, just to understand their desired features
- agents
 - could in principle be *reused* under a variety of situations
 - have *control* over their own *state*
 - are *active*
 - they cannot be invoked
 - agent's control is encapsulated



Focus on. . .

6

Autonomous Systems

- Premises
- What is Autonomy?
- **Technology for Autonomy: Agents & MAS**
- Intentional Agents
- Challenges of Autonomous Systems

7

Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- Challenges



About...

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - **Agents & MAS**
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Agents as Autonomous (Software) Components I

Computational Autonomy

- agents are autonomous as they **encapsulate** (the thread of) **control**
- control does not pass through agent boundaries
 - only data (knowledge, information) crosses agent boundaries
- agents have no interface, cannot be controlled, nor can they be invoked
- looking at agents, **multi-agent systems** (MAS) can be conceived as an aggregation of multiple distinct *loci* of control interacting with each other by exchanging information



Agents as Autonomous (Software) Components II

Agents [Omicini et al., 2008]

Agents are *autonomous computational entities*

genus agents are computational entities

differentia agents are autonomous, in that they encapsulate control along with a criterion to govern it

Agents are *autonomous*

- *autonomy* as the only fundamental and defining feature of agents
- from autonomy, many other features stem
 - autonomous agents *are* interactive, social, proactive, and situated
 - they *might* have goals or tasks, or be reactive, intelligent, mobile
 - they live within MAS, and *interact* with other agents through *communication actions*, and with the environment with *pragmatical actions*

About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - **Agents in JADE**
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



What is JADE? I

- JADE [Bellifemine et al., 2007] stands for *Java Agent DEvelopment Framework*

<http://jade.tilab.com/>

- JADE is a Java-based framework to develop agent-based applications in compliance with the *FIPA specifications* for interoperable, intelligent, multi-agent systems
- FIPA stands for *Foundation for Intelligent Physical Agents*

<http://www.fipa.org/>

- FIPA is the IEEE Computer Society standards organisation that promotes agent-based technology and the interoperability of its standards with other technologies



What is JADE? II

JADE goals

As an **agent-oriented middleware**, JADE pursues the twofold goal of being

- a full-fledged FIPA-compliant *agent platform*
 - hence, it takes charge of all those application-independent aspects – such as agent lifecycle management, communication, distribution transparency, etc. – necessary to develop a MAS
- a simple yet comprehensive *agent development framework*
 - therefore, it provides Java developers a set of APIs to build their own customisations



JADE Main Ingredients

Java

- being fully implemented in Java, JADE is a notable example of a **distributed, object-based, agent-oriented** infrastructure
- hence an interesting example about how to face a design/programming **paradigm shift**

FIPA

- being compliant to FIPA standards, JADE is a *complete* and *coherent* agent platform providing all the necessary facilities to **deploy** MAS
- promoting **interoperability**

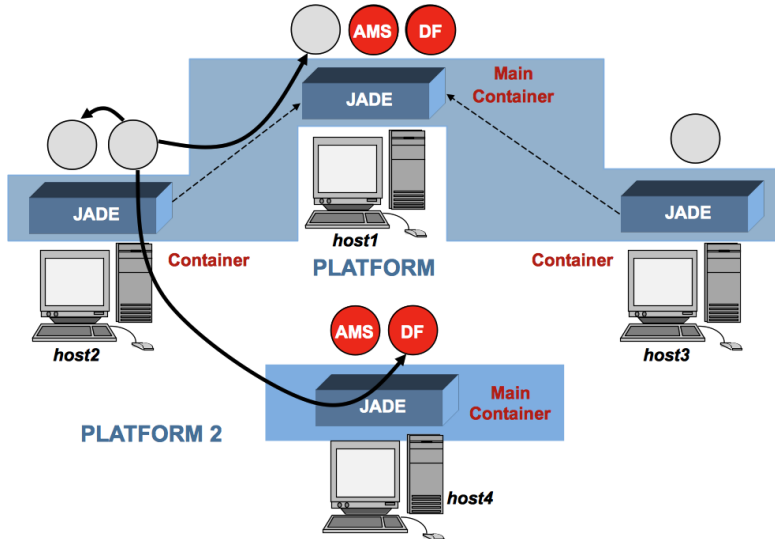
JADE Main Features

JADE offers...

- a **distributed agent platform**, where *distributed* means that a single (logical) JADE system can be split among different networked hosts
- transparent, distributed **message passing** service
- transparent, distributed **naming** service
- white pages & yellow pages **discovering** facilities
- intra-platform agent **mobility** (code & context, to some extent)
- debugging & **monitoring** graphical tools
- ... and much more



JADE Architecture Overview



Focus on. . .

6 Autonomous Systems

- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- **Intentional Agents**
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- Challenges



About. . .

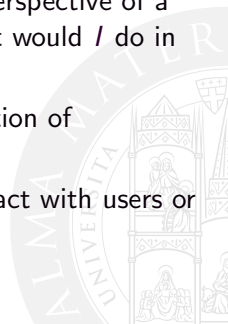
- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - **Agents with Mental States**
 - Agent Reasoning
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Modelling Agents with Mental States I

Modelling agents based on **mental states**...

- eases the development of agents exhibiting complex behaviour
- provides us with a familiar, non-technical way of understanding and explaining agents
- allows the developer to build MAS by adopting the perspective of a cognitive entity engaged in complex tasks—e.g., what would *I* do in the same situation?
- simplifies the construction, maintenance, and verification of agent-based applications
- is useful when the agent has to communicate and interact with users or other system entities



Modelling Agents with Mental States II

The intentional stance [Dennett, 2007]

The intentional stance is the strategy of interpreting the behaviour of an entity (person, animal, artifact, whatever) by treating it as if it were a rational agent who governed its 'choice' of 'action' by a 'consideration' of its 'beliefs' and 'desires'.

*The scare-quotes around all these terms draw attention to the fact that some of their standard connotations may be set aside in the interests of exploiting their central features: **their role in practical reasoning**, and hence in the prediction of the behaviour of practical reasoners.*



Modelling Agents with Mental States III

Agents with mental states

- agents governing their behaviour based on *internal states* that mimic cognitive (human) *mental states*
 - epistemic states** representing agents *knowledge*—their knowledge on the world
 - i.e., percepts, beliefs
 - motivational states** representing agents *objectives*—what they aim to achieve
 - i.e., goals, desires
- the process of selecting one action to execute among the many available based on the actual mental states is called *practical reasoning*

About. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - **Agent Reasoning**
 - BDI Agents
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



Practical vs. Epistemic Reasoning

practical reasoning is reasoning directed towards actions—the process of figuring out what to do in order to achieve what is desired

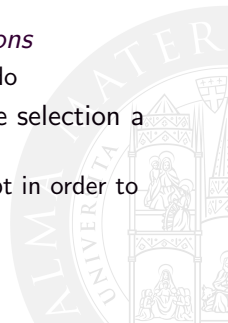
[Bratman, 1987]

Practical reasoning is a matter of weighing conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires/values/cares about and what the agent believes.

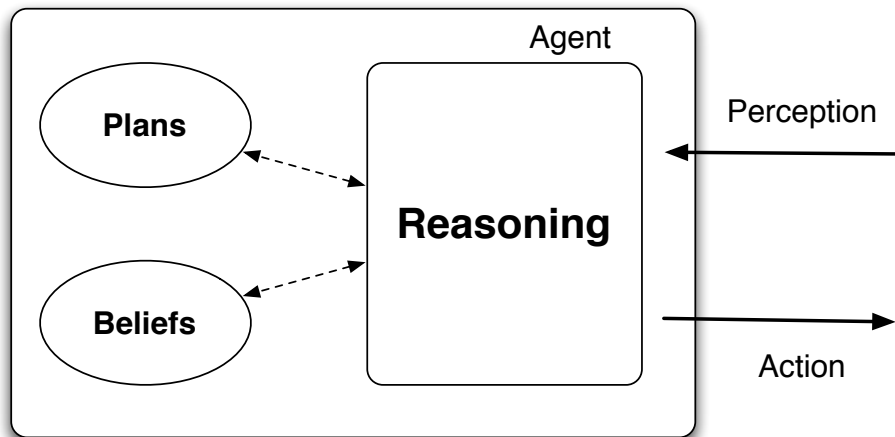
epistemic reasoning is reasoning directed towards knowledge—the process of updating information, replacing old information (no longer consistent with the world state) with new information

Practical Reasoning

- practical reasoning consists of two main cognitive activities
 - **deliberation** when the agent makes decision on what state of affairs the agent desire to achieve
 - **means-ends reasoning** when the agent makes decisions on how to achieve these state of affairs
- the outcome of the deliberation phase are the *intentions*
 - what agent desires to achieve, or what he desires to do
- the outcome of the means-ends reasoning phase is the selection a given course of actions
 - the workflow of the actions the agent intends to adopt in order to achieve its own goals expressed as intentions

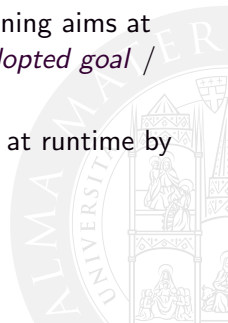


Basic Architecture of a Mentalistic Agent

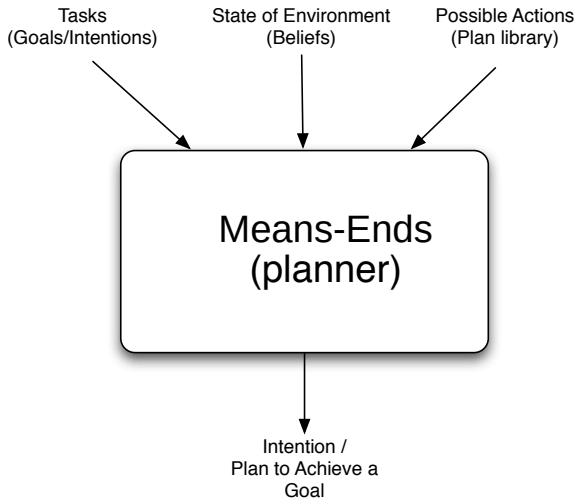


Means-Ends Reasoning I

- the basic idea is to provide agents with three sorts of **representations**
 - representation of **goal** / intention to achieve
 - representation of **actions** / **plans** – in repertoire
 - representation of the **environment**
- given the environmental conditions, means-ends reasoning aims at *devising out a plan* that could possibly *achieve the adopted goal* / intention
- the selected intention is an **emergent property**, reified at runtime by selecting a given plan for achieving a given goal



Means-Ends Reasoning II



About. . .

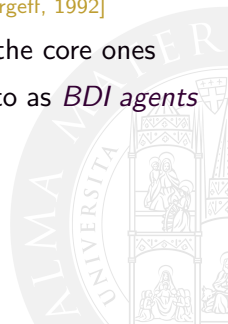
- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Autonomy in Language
 - Autonomy in Biology
 - Autonomy in Philosophy
 - Autonomy in Military Systems
 - Autonomy in Cars
 - Autonomy in Social Sciences & AI
 - Autonomy in Programming Languages
 - Technology for Autonomy: Agents & MAS
 - Agents & MAS
 - Agents in JADE
 - Intentional Agents
 - Agents with Mental States
 - Agent Reasoning
 - **BDI Agents**
 - Challenges of Autonomous Systems
- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



The BDI Framework I

According to [Dasgupta and Ghose, 2011]

- one of the most popular and successful framework for agent technology is defined by Rao and Georgeff [Rao and Georgeff, 1992]
- there, the notions of *belief*, *desire*, and *intention* are the core ones
- hence, agents in this framework are typically referred to as *BDI agents*



The BDI Framework II

beliefs represent at any time the agent's current *knowledge about the world*, including

- information about the current state of the environment inferred from perception devices
- messages from other agents
- internal information

desires represent a *state of the world* the agent is trying *to achieve*

intentions are the chosen means to achieve the agent's desires, and are generally implemented as plans and post-conditions

- as in general it may have *multiple desires*, an agent can have *a number of intentions active at any one time*
- these intentions may be thought of as *running concurrently*, with *one chosen intention active at any one time*

The BDI Framework III

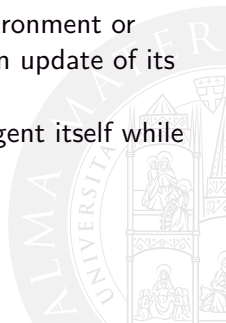
Besides these components, the BDI model includes

plan library — namely, a set of “recipes” representing the *procedural knowledge* of the agent

event queue — where

- *events* — either perceived from the environment or generated by the agent itself to notify an update of its belief base
- *internal subgoals* — generated by the agent itself while trying to achieve a desire

are stored.



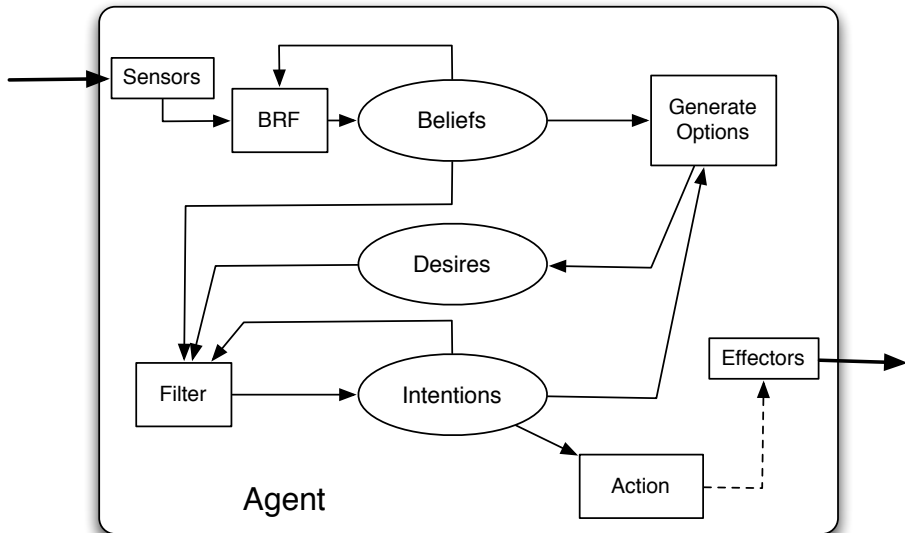
The BDI Framework IV

Plans & plan library

- usually, BDI-style agents do not adopt first principles planning at all
- all plans must be *generated* by the agent programmer *at design time*, which are then *selected* for execution at run time
- pre-programmed plans are collected in the plan library
- the planning done by agents consists entirely of context-sensitive subgoal expansion, which is deferred until a subgoal is selected for execution



Basic Architecture of a BDI Agent [Wooldridge, 2002]



BDI Agents Programming Platforms

Jason (Brasil) <http://jason.sourceforge.net/>
Agent platform and language for BDI agents based on AgentSpeak(L)

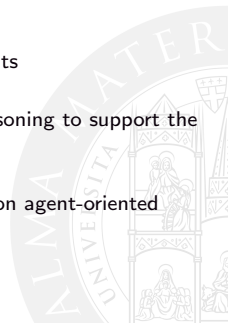
JADEX (Germany) <http://www.activecomponents.org/>
Agent platform for BDI and Goal-Directed Agents

2APL (Netherlands) <http://www.cs.uu.nl/2apl/>
Agent platform providing programming constructs to implement cognitive agents based on the BDI architecture

3APL (Netherlands) <http://www.cs.uu.nl/3apl/>
A programming language for implementing cognitive agents

PRACTIONIST (Italy) <http://practionist.eng.it/>
Framework built on the Bratman's theory of practical reasoning to support the development of BDI agents

ASTRA <http://astralanguage.com/>
A distributed / concurrent programming language based on agent-oriented programming



Focus on. . .

- 6 Autonomous Systems
 - Premises
 - What is Autonomy?
 - Technology for Autonomy: Agents & MAS
 - Intentional Agents
 - **Challenges of Autonomous Systems**

- 7 Conversational Systems
 - Premises
 - Conversation
 - History of Conversational Systems
 - Directions
 - Challenges



A World without Jobs I

Relieving humans from fatigue

- autonomous systems could make life *easier* for every human
 - all low-skilled (and, no-satisfaction) jobs will be delegated to machines
 - working will no longer be a *necessity* for humans
 - *time* will be available for any human to pursue his/her own interests
- making machines work for us will make us globally *richer*
 - welfare measures such as *universal basic income* will become in the end *affordable* for the collectivity



A World without Jobs II

Relieving humans from ageing issues

- humans grow older and older
- progress in science and technology extends our life expectation over any previously known limit
- systems with diverse levels of autonomy (e.g., medical exoskeletons) could allow humans to live even longer and happier lives
- in autonomy, possibly with no fear of solitude



A World without Jobs III

Missing social role

- their *social role* is essential for humans
- it represents the foundation of our self-esteem and social relationships
- till now, our role in the society has mostly been associated to our job
- which role will we play in a *jobless society*?
- what will we do in our *jobless time*?



A World without Jobs IV

The perils of the change

- e.g., autonomous vehicles will soon take the place of *millions* of human drivers everywhere in the world
- in the end, this is ok, but: what will those millions of drivers do of their life?
- what about the possible decades of *unemployment* and *despair* before we can really afford our forthcoming jobless society?



Responsibility without Humans I

Streets without car crashes

- humans are not really good at driving
- most wars in human history have taken a lighter toll on human lives than our streets do every year
- streets and highways without human drivers will soon (in some decades) see very few car crashes, no loss of human life
- possibly, less traffic and better public service, too



Responsibility without Humans II

Wars without human soldiers

- *autonomous weapons* – or, AWS, autonomous weapon systems
[Sartor and Omicini, 2016] – do not require humans on the field
- wars might no longer involve humans
- also, the distinction between defence and attack can be embedded in autonomous weapons



Responsibility without Humans III

Who is to blame?

- whenever machines are in charge, who is responsible of when things go south?
- when machine learning is involved, what is the responsibility of the user?
- will laws and norms evolve in time for ruling the world of AI and autonomous systems?



Responsibility without Humans IV

Machine ethics

- *responsibility* is both a legal and an ethical issue
- as clearly understandable in the case of AWS, where questions are
 - who pulls the trigger?
 - or, better, who decides to pull the trigger?
 - and, based on what evidence?
- **machine ethics** has to be developed, as both a philosophical notion, and a computer science and engineering discipline



Next in Line...

- 6 Autonomous Systems
- 7 Conversational Systems



Focus on. . .

6 Autonomous Systems

- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- **Premises**
- Conversation
- History of Conversational Systems
- Directions
- Challenges



Which Sorts of Systems?

Socio-technical systems (STS) [Whitworth, 2006]

- artificial systems where both *humans* and *artificial* components play the role of system components
 - from online reservation systems to social networks
- most of nowadays systems are just STS
 - or, at least, cannot be engineering and successfully put to work without a proper socio-technical perspective in the engineering stage



Conversational Informatics as a New Paradigm

Conversation for intelligent systems

- **conversation** as a primary means for building *intelligent systems*
- ! STS could fulfil their potential only by providing full conversational capabilities with humans

Conversation with autonomous systems

- conversation as a means for interacting with *autonomous systems*
- ! STS as a conversation between humans and autonomous systems



Focus on. . .

6 Autonomous Systems

- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- **Conversation**
- History of Conversational Systems
- Directions
- Challenges



Conversation for Humans I

Conversation is easy for humans

- **conversation** (in any form) is apparently the most important social activity for humans—more general, for social animals
- humans engage themselves in conversations using all their skill sets without apparent effort
- however, this has not produced a single, comprehensive framework for investigating conversation



Conversation for Humans II

Viewpoints

- verbal communication
- nonverbal communication
- social discourse
- narratives and content flow
- cognitive processes



Stories and Narratives

Storytelling

- conversations as exchanges of small information items that build large stories
- pieces of conversations are brought in and out, back and forth from conversations to individual lives
- to understand facts, to revision beliefs, to interpret and predict events, to build new stories
- **storytelling** has been pointed out as the source of human intelligence: not just to communicate, but to recognise and understand the world
- storytelling plays an essential role in the cognition process



Conversation in a Social Discourse I

Social mechanisms for conversation

- conversation might be either *focussed* or *unfocussed*
- typical conversations have *roles*—e.g., speaker, addressee, side participants, bystanders, eavesdroppers
- conversations have *spatial arrangements* for participants
- the *engagement* in a conversation may vary—for different participant at a given time, for the same participant over time



Conversation in a Social Discourse II

Speech act theory

- speech act theory analyses **utterances** of conversation with respect to social implications
 - locutionary acts** — as the actual action of making an utterance
 - illocutionary acts** as the social acts brought about by given utterances
 - perlocutionary acts** — as the social effect caused by the given utterances
- since it focuses more on the individual communication acts in isolation, speech act theory is precious yet limited in scope and results



Conversation in a Social Discourse III

Joint activity theory

- focus on *interactive aspects*, such as collaborations, coordination among participants, which should be of primary importance in conversation
- basic hypothesis
 - 1 language is fundamentally used for social purposes
 - 2 language use is a species of *joint action*
 - 3 language use always involves speaker's *meaning* and addressee's *understanding*
 - 4 the basic setting for language use is face-to-face conversation
 - 5 language use often has *more than one layer* of activity
 - 6 the study of language use is both a *cognitive* and a *social* science

Making Sense of a Conversation with Multiple Modalities

Using body to talk

- in a conversation, the speaker usually *coordinates* multiple parts of his/her *body*, including his or her eyes, face, hands, head, and torso
- the speaker does this to *express* what he/she wants to communicate
 - e.g., pointing to refer to a portion of the surrounding environment
- complementing or even substituting utterance
- integrating all the diverse forms of communication allow humans to make sense out of a conversation



Cognitive Process I

Modelling the cognitive process underlying conversations

- early attempts in cognitive linguistics, focussing mostly on the cognitive apparatus required to interpret verbal items in a conversation
- *mind reading* is pointed out as an essential in conversations, and more generally in social life
- as the ability to understand motivations, beliefs, goals of the other participants while they speak—*theory of mind*
- ! some other primates have exhibited some theory of mind

Cognitive Process II

Emotions

- **emotion** plays an important role in human intelligence and hence in communication
- different frameworks for emotion classification and composition were developed and proposed, and their effects on conversation discussed and exemplified
- in particular, investigating how emotions relates to goal-oriented activities—in particular, how they may affect goal-oriented conversation



Focus on. . .

6 Autonomous Systems

- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- **History of Conversational Systems**
- Directions
- Challenges



Lines of Research for Conversational Systems I

Early history

- in the 1960s, work on **text-based natural language dialogue systems** begins
 - the first computer programs capable of answering questions typed in English appear
 - simple yet ground-breaking
- in the 1980s, **speech recognition systems** were first developed, towards a more natural user interface were developed
- in the 1990s, **multimodal dialogue systems** allowed users to combine speech with nonverbal input – e.g., gestures, gazes – to interact with a computational system
- in 1987, Apple's Knowledge Navigator previewed the role of AI within an embodied conversational agent

Lines of Research for Conversational Systems II

Understanding and generating stories

- other AI approaches focussed on how narratives and stories could be understood and generated
- towards *transactional systems*
- such approaches are critical to address content management of conversational systems
- not just telling stories
- but also learning from conversations, and retaining (pieces of) accumulated conversations in memory, for future (re)use

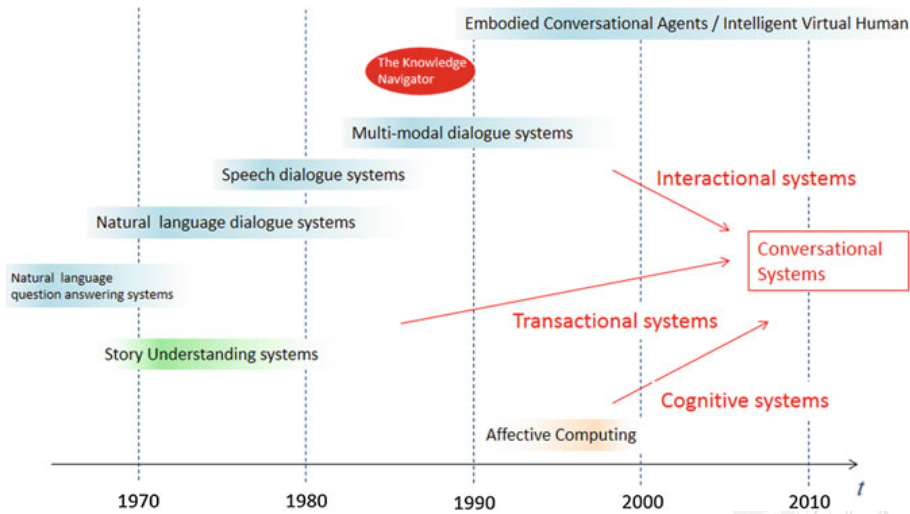
Lines of Research for Conversational Systems III

Cognitive systems

- since the 1990s, focus on cognitive system exhibiting more human-like intellectual behaviours resulting from autobiographical dynamic memory, emotion and theory of mind
- among others, *affective computing* became a visible line of research



Lines of Research for Conversational Systems IV



[Nishida et al., 2014]

Lines of Research for Conversational Systems V

Integration—as usual

It can be easily observed that

- without proficient interaction, stories may be told poorly
- without stories, conversation is boring
- without a cognitive model, we cannot build an attractive conversational agent

As a result, **integration** is considered necessary for conversational systems to be successful



Early Examples of Natural Language Dialogue Systems I

Baseball (1961)

- natural language question answering system built to answer questions about baseball games
- based on a small dictionary and programmed grammar
- produces database retrieval commands to answer information requests in natural language (English)



Early Examples of Natural Language Dialogue Systems II

LUNAR (1973)

- Lunar Natural Sciences Natural Language Information System (LUNAR)
- a natural language question answering system that can answer queries about the rock samples brought from the moon, featuring
 - *syntactic analysis* using heuristic and semantic information to select the most likely syntactic analysis from candidates
 - (domain specific) *semantic representation* used as an intermediate representation to produce database queries
 - Augmented Transition Network Grammar (ATNG) to analyse inputs by combining syntactic and semantic analysis
- able to answer complex queries such as

Give me all lunar samples with Magnetite

In which samples has Apatite been identified

How many samples contain Titanium

Which rocks do not contain Chromite and Ulvospinel.

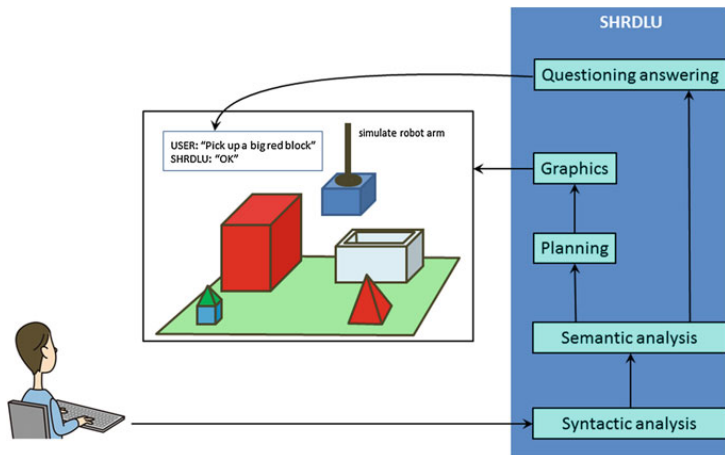
Early Examples of Natural Language Dialogue Systems III

SHRDLU (1972)

- interacting program able to understand the meaning of natural language expressions with respect to the *block world*
- while limited in scope, first approach prototyping the *entire process of natural language understanding* in computational terms
- combining all required components (natural language processing, planner, ...) in a virtual agent behaving as if it understands the meaning of natural language
- e.g., given the utterance “pick up a big red block”, SHRDLU can determine which physical object is referred to and what goal the user wants it to achieve



Early Examples of Natural Language Dialogue Systems IV



[Nishida et al., 2014]

Early Examples of Natural Language Dialogue Systems V

ELIZA (1966)

- no extensive grammar, no semantic knowledge—just
 - identification of keywords
 - detection of a minimal context
 - selection of proper transformations
 - response generation when no keywords are found in the input
- basically, cunning pattern matching—but funny, and brought some fans to AI at the time



Speech Dialogue Systems and Multimodal Interfaces I

HEARSAY-II (1980)

- HEARSAY-II allowed the user to actually **speak** to the system
- based on a *blackboard architecture*
 - a shared database (blackboard)
 - a collection of cooperating processes
- information representation at seven levels of abstraction, from the parameter level to the phrase level
- knowledge sources processing incoming signal concurrently, generating hypothesis for interpretation at the different levels of abstractions, that should converge in a single coherent interpretation
- changes handled by the *blackboard monitor*
- later replaced by the goal blackboard and goal processor, dedicated to *goal-oriented processing*

Speech Dialogue Systems and Multimodal Interfaces II

Put-That-There (1980)

- speech dialogue systems extended to multimodal interfaces
- so as to use more than one modality of communication in human-computer interaction
- Put-That-There allowed the user to arrange simple shapes on a large graphics display surface by *voice* and a simultaneous *pointing gesture*



Speech Dialogue Systems and Multimodal Interfaces III

Multimodal interaction systems

- *fusion* as the key issue
- robust and efficient algorithms to interpret multimodal input from heterogeneous sensors by handling errors and missing data
- to produce multimodal output, output from different knowledge sources are tailored and synchronised
- this involves content selection and organisation, coordinated distribution of information on available modalities, modality-specific content realisation, and laying out generation results
- *presentation* may be characterised as a knowledge-based, goal-directed activity under constraints
- planning algorithm to assemble the components into a coherent structure of presentation

Embodied Conversational Agents & Intelligent Virtual Humans I

The Knowledge Navigator (1987) & Peedy (1997)

- The Knowledge Navigator video (TKNV) pioneered the idea of a visual human-like embodied conversational agent capable of multimodal conversational interaction
- anthropomorphism is debatable, but surely can allow the users to apply proficient inter-human communication skills
- Peedy, the Conversational Personal Assistant, is a conversational system in which a parrot-agent named Peedy helps the user select songs from a collection of audio CDs—and embodied the key concepts introduced in TKNV

Embodied Conversational Agents & Intelligent Virtual Humans II

Rea (1999)

- Rea is an embodied conversational agent that can effectively use a human-like body for verbal and nonverbal communication
- the system actually implements conversational interactions, in contrast to employing conversation as a metaphor of interface
- focused on effective use of nonverbal communication media as social signals, such as gaze or hand gestures, to achieve communication functions, such as acknowledgment of the communication partner or taking turns
- implemented the ideas in the real-estate domain, where Rea was characterised as a real-estate agent who shows users the features of various houses
- full symmetry between input and output modalities were intended to allow Rea to participate in human-computer conversation with equal footing
- ! KQML (standard markup / agent communication language) was employed at the system level so that heterogeneous modules could interact with each other without constraints by exchanging messages in a common representation language

Embodied Conversational Agents & Intelligent Virtual Humans III

Methodological progress after Rea

- specific **script/markup languages** for specifying the behaviour of embodied conversational agents—e.g., AIML, Microsoft Agent, MPML, STEP, BML
- **corpus-based generation of behaviours** in which the agent's behaviour is generated in two phases
 - first one or more corpus is built by collecting and annotating a sufficient amount of observational data for the target communicative behaviours
 - then, parameter values are determined based on the analysis of the data accumulated in the corpora



Story Understanding/Generation Systems

Knowledge to understand stories

- after early experiences, it was soon clear that some **background knowledge** (structure) was required in order to understand stories
- e.g., *scripts* to represent knowledge of stereotypical scenes
- or, *plans* to understand goal-oriented stories
- theory of *dynamic memory* was used to learn from stories



Cognitive Computing I

Believability and lifelikeness of interacting agents

- interacting agents should be believable in order to gain sufficient sense of presence and reality
- a **believable agent** can provide the illusion of life and permit the audience's suspension of disbelief
- e.g., in the Oz project (1994), virtual worlds are inhabited by interactive characters within which the user experiences a story from a first person perspective
- e.g., The Artificial Life Interactive Video Environment (ALIVE) system allows the user to engage in entertainment interaction with an animal-like agent

Cognitive Computing II

Affective computing

- *emotion* plays a critical role in communication and decision making
- e.g., an emotionless machine intelligence might be vulnerable; it may not recognise a danger that is not logically deducible, which even a young child could easily identify
- in communication, the ability to express and interpret emotion is mandatory for an intelligent agent to engage *empathically*
- according to Picard (1997), *emotional intelligence* consists in the core of human intelligence, argued that computational intelligence needs the ability to recognise and express emotions in order to be a genuinely intelligent partner
- Affective Reasoner (1992) used the OCC model to realise emotional natural language dialogues

Focus on. . .

6 Autonomous Systems

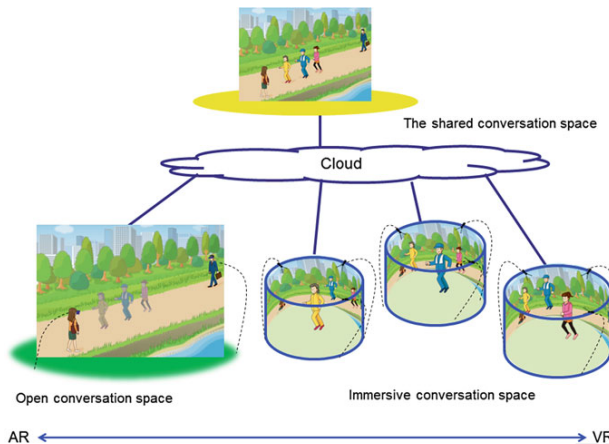
- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- **Directions**
- Challenges



Smart Conversation Space I



[Nishida et al., 2014]

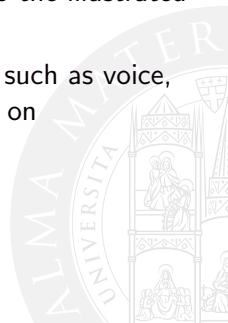
Smart Conversation Space II

- open conversation space enhanced by augmented reality (AR) where people can move around talking to each other with reference to objects and events in the space
- an immersive conversation space where each participant is completely encircled by large screens and the behaviour of the participant is sensed and reflected on his/her image in the shared conversation space



Immersive Collaborative Interaction Environment I

- an immersive conversation space uses a series of large screens encircling the user to provide rich ambient audio-visual information so that s/he may be involved and feel closely attached to the illustrated scene
- the user in return responds with diverse social signals such as voice, body motions, altered gaze, facial expressions, and so on



Focus on. . .

6 Autonomous Systems

- Premises
- What is Autonomy?
- Technology for Autonomy: Agents & MAS
- Intentional Agents
- Challenges of Autonomous Systems

7 Conversational Systems

- Premises
- Conversation
- History of Conversational Systems
- Directions
- **Challenges**



Dangers I

The dangers of AI?

- it looks quite cool these days to chit-chat about the dangers of AI and autonomous machines: robots will take the power (steal our jobs, win our women / men, kill everybody), *yeah*, soooooon.
- however, the perils of autonomous systems are not the same posed by conversational informatics at its peak



Dangers II

Sociality vs. reasoning

- driven by our troublesome experience with social networks, everybody knows about *cognitive biases* such as confirmation bias and Dunning–Kruger effect
 - basically, we do not easily let facts stand in our way
- we are hyper-social animals: “We never think alone”
- reasoning evolved *after* our ability to interact socially
- we privilege social interaction over reasoning
- and, *looking right in front of our group* is more important than *being really right*—whatever this means
 - the very structure of our judiciary process should tell us a lot about this



Dangers III

Dangers of conversational informatics

- when social interaction becomes human interaction within STS the same issues may occur
- in particular when, for the info-plosion, we are overwhelmed by data, information, knowledge—*facts* of some sort, difficult to be understood and interpreted to compose the overall picture
- what happens when we do no longer interact with some database or web page to look for some facts, but we engage instead in a conversation with a conversational agent?
- we are probably going to *trust* the conversational agent by default—knowledge acquired conversationally is trusted at a deeper level than any other knowledge
- e.g., fabricating false evidence is nowadays so easy, making it accepted as real could be much easier via conversational systems



Dangers IV

What could we do?

- the point is that *info-* and *techno-plosion* changed / are changing our surrounding environment into the knowledge-intensive, hyper-social environment we nowadays live in *too fast* for us to evolve and be ready for it
- of course, the first measure regard the way in which we *build* those systems that could endanger us—at least, for the aspects we really can control
- so, for instance, one simple advice would be *not* to build conversational agents like us, and make them instead ready – their conversational intelligence, first of all – for *rational interaction*
- e.g., empathy is ok in order to better understand humans along with their motivations, but should not overcome rationality

Part III

Bibliography



References I



Barber, D. (2012).

Bayesian Reasoning and Machine Learning.
Cambridge University Press.



Bellifemine, F. L., Caire, G., and Greenwood, D. (2007).

Developing Multi-Agent Systems with JADE.
Wiley.



Bratman, M. E. (1987).

Intention, Plans, and Practical Reason.
Harvard University Press.



Brooks, R. A. (1986).

Achieving artificial intelligence through building robots.
Technical report, Massachusetts Institute of Technology (MIT).



Brooks, R. A. (1991a).

Intelligence without reason.
In Mylopoulos, J. and Reiter, R., editors, *12th International Joint Conference on Artificial Intelligence (IJCAI 1991)*,
volume 1, pages 569–595, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.



Brooks, R. A. (1991b).

Intelligence without representation.
Artificial Intelligence, 47:139–159.



Campbell, M., Hoane, A., and Hsu, F.-h. (2002).

Deep Blue.
Artificial Intelligence, 134(1):57–83.



References II



Castelfranchi, C. (1995).

Guarantees for autonomy in cognitive agent architecture.

In Wooldridge, M. J. and Jennings, N. R., editors, *Intelligent Agents*, volume 890 of *Lecture Notes in Computer Science*, pages 56–70. Springer Berlin Heidelberg.

ECAI-94 Workshop on Agent Theories, Architectures, and Languages (ATAL) Amsterdam, The Netherlands 8–9 August 1994. Proceedings.



Colmerauer, A. and Roussel, P. (1996).

The birth of Prolog.

In Bergin Jr., T. J. and Gibson Jr., R. G., editors, *History of programming languages—II*, volume II, pages 331–367. ACM, New York, NY, USA.



Conte, R. and Castelfranchi, C. (1995).

Cognitive and Social Action.

UCL Press.



Dasgupta, A. and Ghose, A. K. (2011).

BDI agents with objectives and preferences.

In Omicini, A., Sardina, S., and Vasconcelos, W., editors, *Declarative Agent Languages and Technologies VIII*, volume 6619 of *Lecture Notes in Computer Science*, pages 22–39. Springer Berlin Heidelberg.



Dennett, D. (2007).

Intentional systems theory.

In Beckermann, A. and Walter, S., editors, *Oxford Handbook of the Philosophy of Mind*, chapter 19. Oxford University Press.

References III



Edwards, G. (2013).

Military autonomous & robotic systems. Considerations for the way forward from a UK military perspective. *Air Power Review*, 16(3):50–71.



Encyclopedia of Cognitive Science (2006).

Encyclopedia of Cognitive Science.
John Wiley & Sons.



Furcy, D. and Koenig, S. (2006).

STRIPS.
In [Encyclopedia of Cognitive Science, 2006].



Lenat, D. B., Prakash, M., and Shepherd, M. (1985).

CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI Magazine*, 6(4).



Lippmann, R. P. (1987).

An introduction to computing with neural nets.
IEEE ASSP Magazine, 4(2):4–22.



Lloyd, J. W. (1984).

Foundations of Logic Programming.
Springer, 1st edition.



McCarthy, J. (1981).

History of LISP.
In Wexelblat, R. L., editor, *History of Programming Languages I*, pages 173–185. ACM, New York, NY, USA.



References IV



Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors (1983).
Machine Learning: An Artificial Intelligence Approach.
Symbolic Computation. Springer.



Newell, A. (1959).
Report on a general problem-solving program.
In *International Conference on Information Processing*, pages 256–264.



Newell, A. and Simon, H. (1956).
The logic theory machine—a complex information processing system.
IRE Transactions on Information Theory, 2(3):61–79.



Nishida, T., Nakazawa, A., Ohmoto, Y., and Mohammad, Y. (2014).
Conversational Informatics. A Data-Intensive Approach with Emphasis on Nonverbal Communication.
Springer Japan, Tokyo.



Odell, J. (2002).
Objects and agents compared.
Journal of Object Technologies, 1(1):41–53.



Omicini, A., Ricci, A., and Viroli, M. (2008).
Artifacts in the A&A meta-model for multi-agent systems.
Autonomous Agents and Multi-Agent Systems, 17(3):432–456.
Special Issue on Foundations, Advanced Topics and Industrial Perspectives of Multi-Agent Systems.



References V



Papadimitriou, C. H. (1977).

The Euclidean travelling salesman problem is NP-complete.
Theoretical Computer Science, 4(3):237–244.



Rao, A. S. and Georgeff, M. P. (1992).

An abstract architecture for rational agents.

In Nebel, B., Rich, C., and Swartout, W. R., editors, *3rd International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pages 439–449, Cambridge, MA, USA. Morgan Kaufmann. Proceedings.



Rosslenbroich, B. (2014).

On the Origin of Autonomy. A New Look at the Major Transitions in Evolution.
History, Philosophy and Theory of the Life Sciences. Springer International Publishing.



SAE J3016 (2018).

Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.
Surface Vehicle Recommended Practice J3016TM, SAE International.
Issued 2014-01-16. Revised 2016-09-30; 2018-06-15.



Sartor, G. and Omicini, A. (2016).

The autonomy of technological systems and responsibilities for their use.

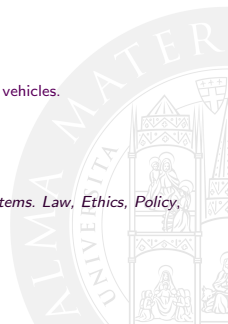
In Bhuta, N., Beck, S., Geiß, R., Liu, H.-Y., and Kreß, C., editors, *Autonomous Weapon Systems. Law, Ethics, Policy*, chapter 3, pages 39–74. Cambridge University Press, Cambridge, UK.



Searle, J. R. (1980).

Minds, brains, and programs.

The Behavioural and Brain Sciences, 3:417–457.



References VI



Shortliffe, E. H. (1976).

Computer-based medical consultations: MYCIN, volume 2 of Artificial Intelligence Series.
North-Holland.



Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016).

Mastering the game of Go with deep neural networks and tree search.
Nature, 529:484–489.



Standage, T. (2002).

The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine.
Walker & Company.



Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L., Koelen, C., Markey, C., Rummel, C., Niekirk, J. v., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., and Mahoney, P. (2006).

Stanley: The robot that won the DARPA Grand Challenge.
Journal of Field Robotics, 23(9):661–692.



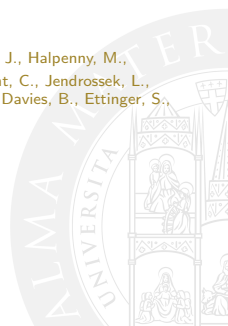
Turing, A. M. (1937).

On computable numbers, with an application to the entscheidungsproblem.
Proceedings of the London Mathematical Society, s2-42(1):230–265.



Turing, A. M. (1950).

I.—computing machinery and intelligence.
Mind, LIX(236):433–460.



References VII



Ward, N. (2006).

SHRDLU.

In [Encyclopedia of Cognitive Science, 2006].



Whitworth, B. (2006).

Socio-technical systems.

In Ghaou, C., editor, *Encyclopedia of Human Computer Interaction*, pages 533–541. IGI Global.



Wooldridge, M. (2018).

Artificial Intelligence, volume 117 of *Ladybird Expert*.

Penguin Random House UK.



Wooldridge, M. J. (2002).

An Introduction to MultiAgent Systems.

John Wiley & Sons Ltd., Chichester, UK.



Intelligent & Autonomous Systems

Digital Culture for Digital Transformation

Andrea Omicini
andrea.omicini@unibo.it

DISI / BBS, Università di Bologna

Evening MBA Program, 16 April 2019

