Artificial Intelligence

From intelligence to rationality?

Viola Schiaffonati
viola.schiaffonati@polimi.it

Can machine think?

The birth of Artificial Intelligence

A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1) Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2) How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning

The dream of machine intelligence

*"The study is to proceed on the basis of the conjecture that **every aspect** of **learning** or any **other feature** of **intelligence** can in principle be so **precisely described** that a machine can be made to **simulate** it."*

(John McCarthy, Marvin L. Minsky,
Nathaniel Rochester,
and Claude E. Shannon 1955)

“Can machines think?”

VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly

"Computing Machinery and Intelligence"

- Turing 1950
- **Difficulties** in the **definition** of the **terms**
  - Machine
  - Think
- **Replacement** of the **question** with another closely related and expressed in **unambiguous words**

The ‘imitation game’

- New form of the problem **described in terms** of a **game** (**imitation game**)
  - Three people: a man (A), a woman (B), an interrogator (C)
  - C is apart from the other two
  - C’s object: to **determine** which of the other two is the man and which is **the woman** by means of questions
  - A’s object: to cause C to make the wrong identification
  - B’s object: to help C

The Turing test

"Can machine think?"

replaced by

"What will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?"

The Turing test and intelligence

- Two **possible interpretations** for the test
  - **Operational definition** of **intelligence** to overcome the problems of a formal definition of intelligence
  - Test able **to detect intelligence**
- **Conventional** definition of intelligence

The problem of intelligence

- **Wide** meaning
  - Evolved brain structure able to solve new problems
  - Symbolic abilities
  - Adaptive relationship with environment
- **Specific** meaning
  - Set of very complex mental processes (only in humans)
  - Logical reasoning, ability to pursue a long term goal, critical capabilities, …

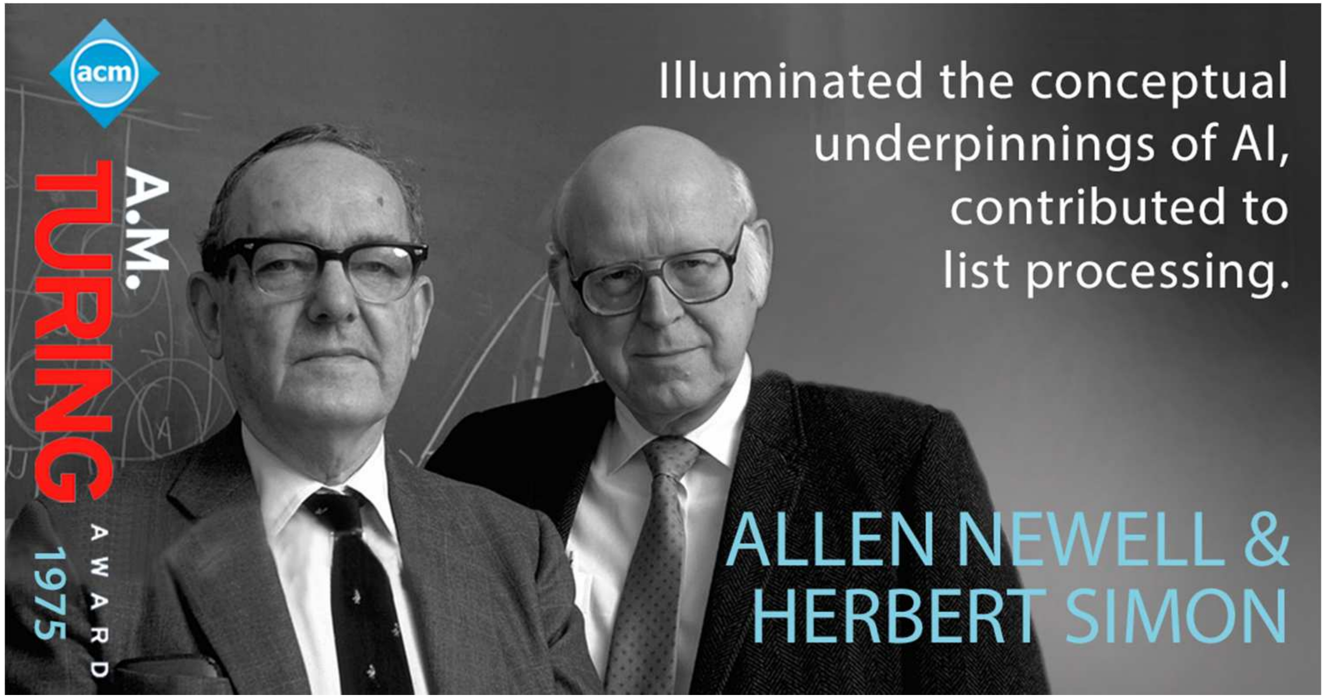
Psychology and intelligence

- **Scarce attention** to the study of human intelligence
  - Objective **difficulty**
  - **Problem of method**: higher mental states cannot be objects of rigorous empirical research
  - Emphasis on the study of **learning** (in particular animal learning)

A different approach

*" A physical symbol system has the **necessary** and **sufficient** means for **general intelligent action**"*

(Newell and Simon 1976)



Symbols and Search

1975 ACM Turing
Award Lecture

The 1975 ACM Turing Award was presented jointly to Allen Newell and Herbert A. Simon at the ACM Annual Conference in Minneapolis, October 20. In introducing the recipients, Bernard A. Galler, Chairman of the Turing Award Committee, read the following citation:

"It is a privilege to be able to present the ACM Turing Award to two friends of long standing, Professors Allen Newell and Herbert A. Simon, both of Carnegie-Mellon University.

"In joint scientific efforts extending over twenty years, initially in collaboration with J.C. Shaw at the RAND Corporation, and subsequently with numerous faculty and student colleagues at Carnegie-Mellon University, they have made basic contributions to artificial intelligence, the psychology of human cognition, and list processing.

"In artificial intelligence, they contributed to the establishment of the field as an area of scientific endeavor, to the development of heuristic programming generally, and of heuristic search, means-ends analysis, and methods of induction, in particular; providing demonstrations of the sufficiency of these mechanisms to solve interesting problems.

"In psychology, they were principal instigators of the idea that human cognition can be described in terms of a symbol system, and they have developed detailed theories for human problem solving, verbal learning and inductive behavior in a number of task domains, using computer programs embodying these theories to simulate the human behavior.

"They were apparently the inventors of list processing, and have been major contributors to both software technology and the development of the concept of the computer as a system of manipulating symbolic structures and not just as a processor of numerical data.

"It is an honor for Professors Newell and Simon to be given this award, but it is also an honor for ACM to be able to add their names to our list of recipients, since by their presence, they will add to the prestige and importance of the ACM Turing Award."

Computer Science as Empirical Inquiry: Symbols and Search

Allen Newell and Herbert A. Simon

Computer science is the study of the phenomena surrounding computers. The founders of this society understood this very well when they called themselves the Association for Computing Machinery. The machine—not just the hardware, but the programmed, living machine—is the organism we study.

This is the tenth Turing Lecture. The nine persons who preceded us on this platform have presented nine different views of computer science. For our organism, the machine, can be studied at many levels and from many sides. We are deeply honored to appear here today and to present yet another view, the one that has permeated the scientific work for which we have been

Key Words and Phrases: symbols, search, science, computer science, empirical, Turing, artificial intelligence, intelligence, list processing, cognition, heuristics, problem solving.

CR Categories: 1.0, 2.1, 3.3, 3.6, 5.7.

The authors' research over the years has been supported in part by the Advanced Research Projects Agency of the Department of Defense (monitored by the Air Force Office of Scientific Research) and in part by the National Institutes of Mental Health.

Authors' address: Carnegie-Mellon University, Pittsburgh.

Physical symbol system

- Newell and Simon 1976
- **Symbols**, **expressions**, **collection** of **processes**
- **Machine** producing through time an evolving collection of symbol structures
- **World** to which symbols and expressions refer

The physical symbol system hypothesis

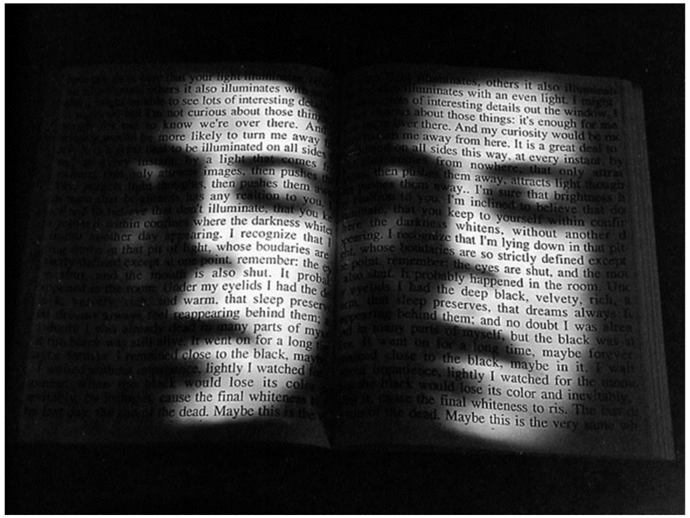
*" A physical symbol system has the **necessary** and **sufficient** means for **general intelligent action**"*

(Newell and Simon 1976)

- **Necessary**: any system exhibiting general intelligence is proved to be a physical symbol system
- **Sufficient**: any physical symbol system can be organized to exhibit general intelligence

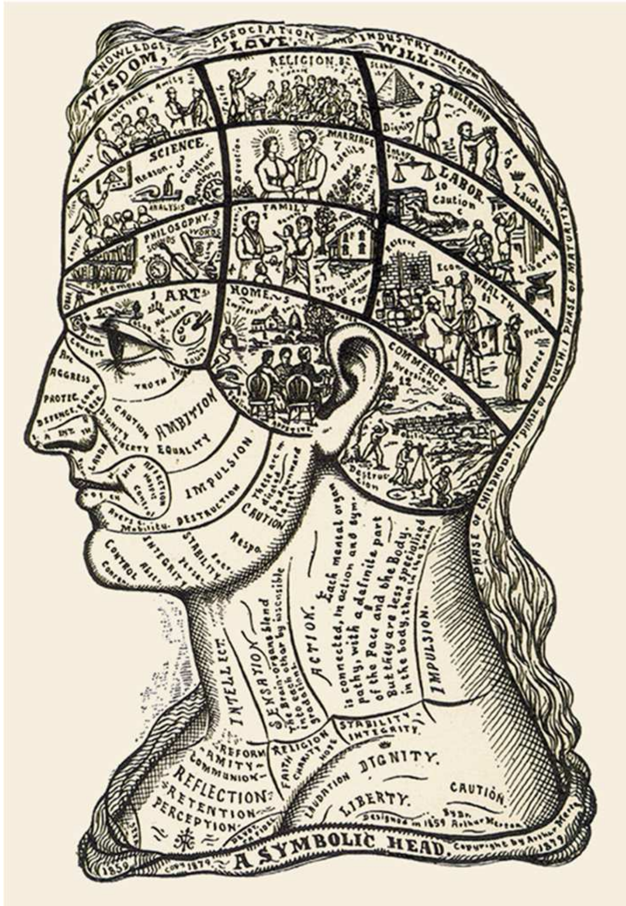
General intelligent action

- The same **scope** of intelligence (as seen in human action)
  - Behavior **appropriate** to the ends of the system
  - Behavior **adaptive** to the demands of the environment



Some consequences

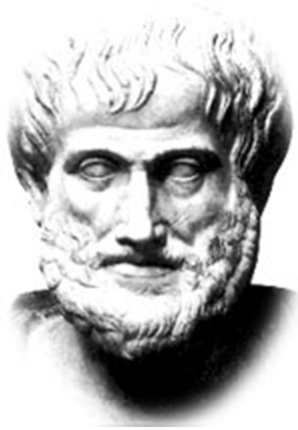
- **Symbolic capacity** as the source of every intelligent behavior
- Explanation of human intelligent behavior in terms of **symbol systems**
- Human intelligent action can be **modeled** by a symbol system



From intelligence to rationality

- Shifting from intelligence to **rationality**
  - History of AI
    - **Impossibility** of **general intelligence**
    - **Context intelligence** (expert systems)
    - The **“whole” agent problem**
  - Rationality as ideal concept of intelligence
    - **Easier** to define criteria for **rationality**

Theoretical and practical rationality



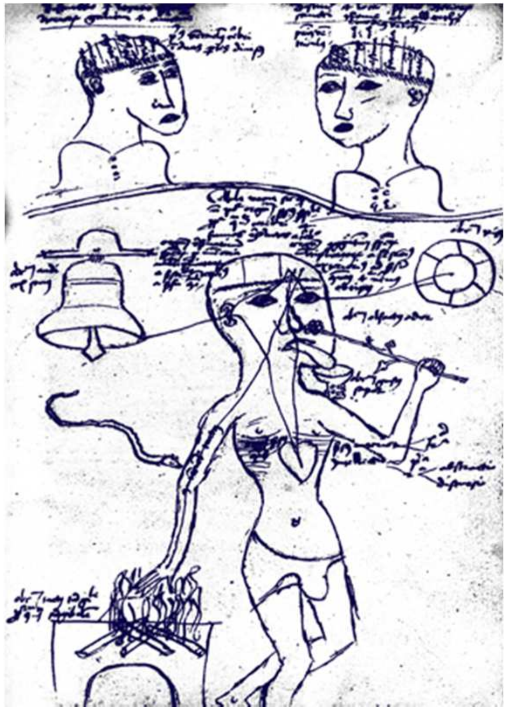
- Different **objects** and **methods** (Aristotle)
- **Theoretical rationality**
  - Object: part of reality independent from humans
  - Method: always true demonstrations
- **Practical rationality**
  - Object: part of reality connected to the human actions
  - Method: mainly true demonstrations

Economic (quantitative) rationality

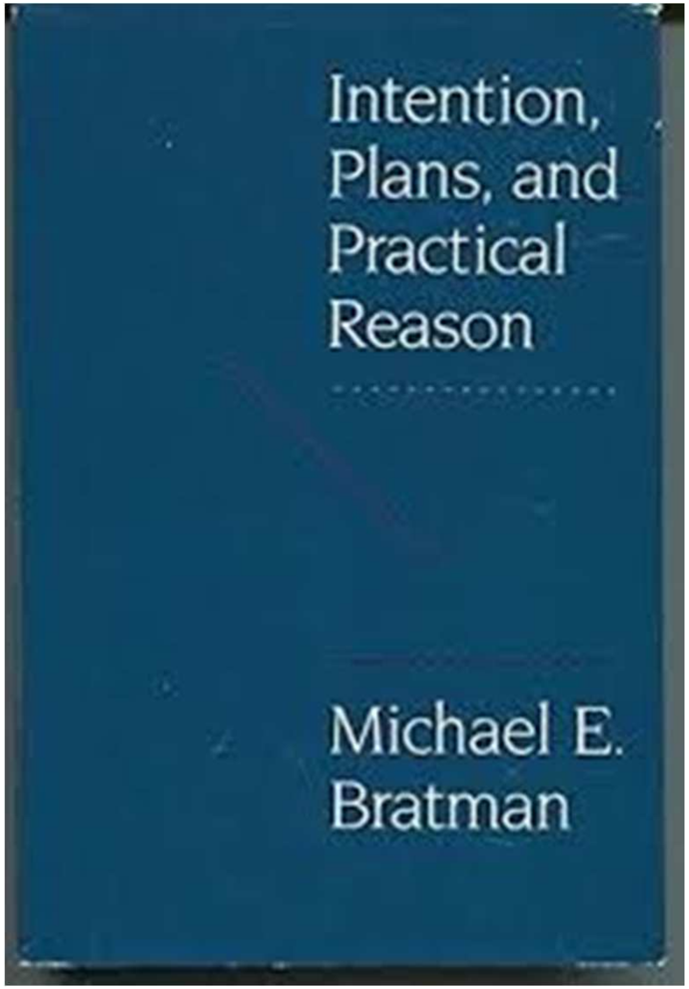
- Choice of **means** and **behaviors** to achieve some **goals**
- **Numeric utility function** to be optimized
- Issues
  - Choice under **uncertainty conditions**
  - **Time** and **cost** to gather information
- Limited rationality (*Simon, Shaw, and Newell 1963*)

The classical model of (practical) rationality

- Rational actions caused by **beliefs** and **desires**
- Rationality as a matter of **obeying rules**
  - Rules to distinguish between rational and irrational thoughts and behaviors
- Rationality as a separate cognitive faculty
- Practical reason as starting from an **agent's primary ends** (goals, desires, objective, and purposes)



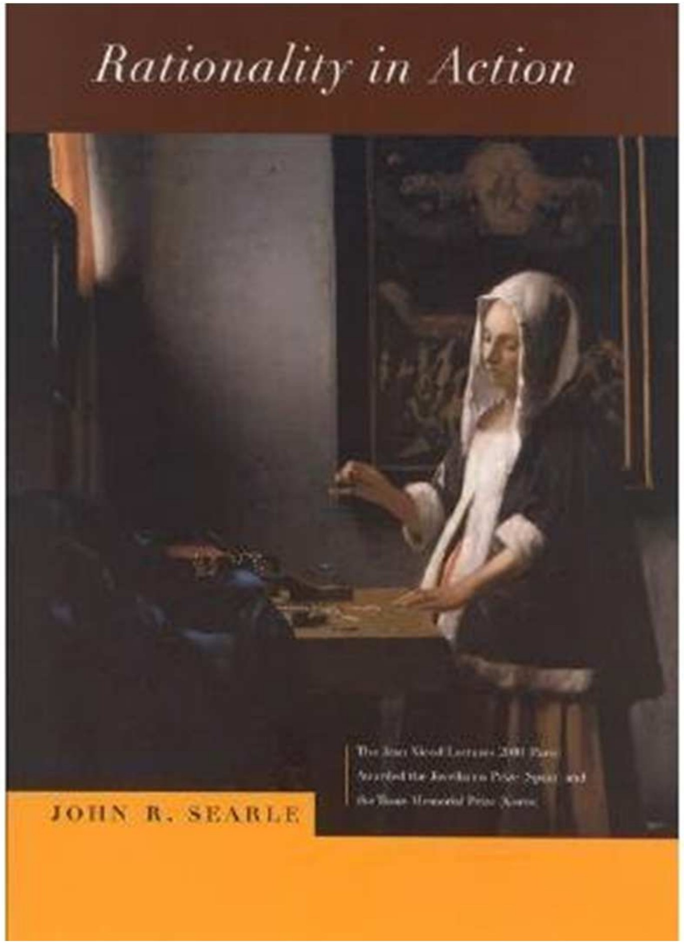
Another model of rationality

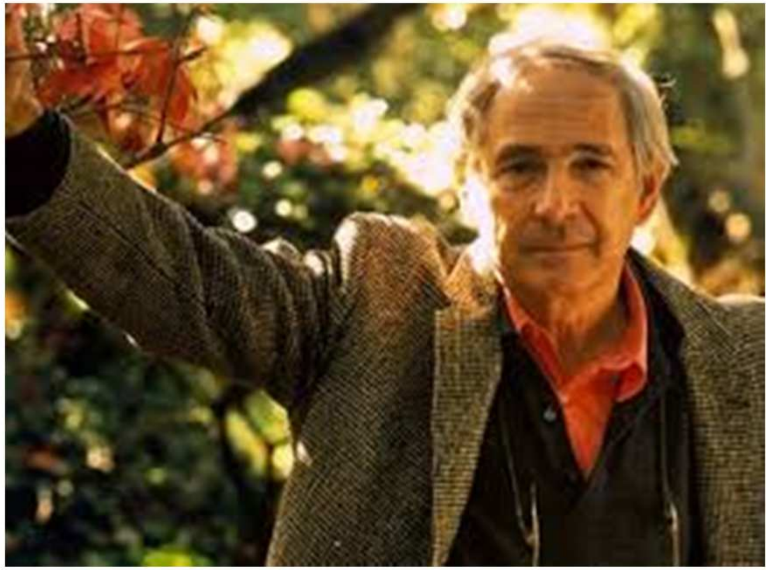


The BDI model of rationality

- Bratman 1987
- **Beliefs**
  - Way of representing the world
- **Desires** (or **goals**)
  - Desired end states
  - Way of remembering where the system wants to get
- **Intentions**
  - Committed plans or procedures
  - Capability of reconsidering the adopted plans

A non causal theory of rationality





Rationality in action

- Searle 2001
- **Reasons not sufficient** to cause **intentional actions**
  - Human behavior not described in terms of laws and causes
  - **Free will**
- Distinction between **prior intention** and **intention-in-action**

Rational (artificial) intelligence

- Rationality as **reasons to act**
  - Doing the right thing or the most successful
- Rationality depends on
  - **Performance measure** (success criterion)
  - Agent’s **prior knowledge** of the environment
  - **Actions** the agent can perform
  - Agent’s **percept sequence**

Rationality and action

*"For each possible percept sequence, a rational agent should select an action that is expected to **maximize** its **performance measure**, given the **evidence** provided by the **percept sequence** and whatever **built-in knowledge** the agent has."*

(Russell and Norvig 2009)

Theoretical worries and practical directions

- **Psychological approach**
  - Observation and experiments on human behavior in task requiring intelligence
  - Programming of symbol systems **to model** the observed **human behavior**
- **Engineering approach**
  - **Conventionally setting** of an **action** or behavior as **intelligent**
  - Designing and realizing a **machine** implementing this behavior

References

- Bratman, M. (1987) *Intention, Plans, and Practical Reason*, CSLI Publications
- Newell, A., Simon, H. (1976) "Computer Science as Empirical Inquiry: Symbols and Search" *Communications of the ACM*, 19:3, 113-126
- Searle, J. (2001) *Rationality in Action*, The MIT Press
- Turing, A. (1950) "Computing Machinery and Intelligence", *Mind* 59, 433-460