# Intelligenza Artificiale

## III Modulo, Curriclum T5

# Alcune informazioni

♢ *Maurizio Gabbrielli*

♢ Libro di testo: *Russel, Norvig. Intelligenza Artificiale: un approccio moderno, vol. I.* (AIMA Berkeley) http://aima.cs.berkeley.edu/

♢ Lucidi anche Russel and Norvig

# Programma

- Cosa è l'intelligenza
- Cosa è l'intelligenza artificiale (IA)
- A. M. Turing e l'IA
- La nozione di agente (intelligente ?)
- Strategie di ricerca (non informata, informata, con avversari).
- Alcuni linguaggi per IA
- Alcune applicazioni di IA

# Outline

- ♢ What is AI?
- ♢ A brief history
- ♢ The state of the art

# What is AI?

| |
|---|
| **Systems that think like humans** |
| **Systems that think rationally** |
| **Systems that act like humans** |
| **Systems that act rationally** |

# Thinking humanly: Cognitive Science

1960s "cognitive revolution": information-processing psychology replace prevailing orthodoxy of behaviorism (this was concerned with "measures" of stimula/answers only)

Requires scientific theories of internal activities of the brain
- What level of abstraction? "Knowledge" or "circuits"?
- How to validate? Requires
  1) Predicting and testing behavior of human subjects (top-down)
  or 2) Direct identification from neurological data (bottom-up)

Both approaches (roughly, Cognitive Science and Cognitive Neuroscience are now distinct from AI

Both share with AI the following characteristic:

**the available theories do not explain (or engender) anything resembling human-level general intelligence**

Hence, all three fields share one principal direction!

# Thinking humanly: imitate the brain

Why not imitate the human brain ? We do not know enough:

– Broca (1861) relation between language and specific areas in the brain
  – Golgi (1873) first visualization of neurons
  – Berger (1929) EEG
  – Ogawa et al (1990) FMRI: measurement of cognitive processes.

Today

– We know some relations between areas in the brain and specific functions
  – do not know how a cognitive process works
  – do not know how memory works

# A comparison

| | Computer | Brain |
|---|---|---|
| Comp. units | $10^8$ logical ports | $10^{11}$ neurons |
| Memory | $10^{10} + 10^{11}$ bit | $10^{11} + 10^{14}$ (neurons + synapses) |
| Clock | $10^{-9}$ sec | $10^{-3}$ sec (parallel !) |
| Bandwidth | $10^{10}$bit/sec | $10^{14}$ bit/sec |

Brain 100.000 times faster than computer in performing activities (all neurons work in parallel).

# Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of logic:
**notation** and **rules of derivation** for thoughts;
may or may not have proceeded to the idea of mechanization

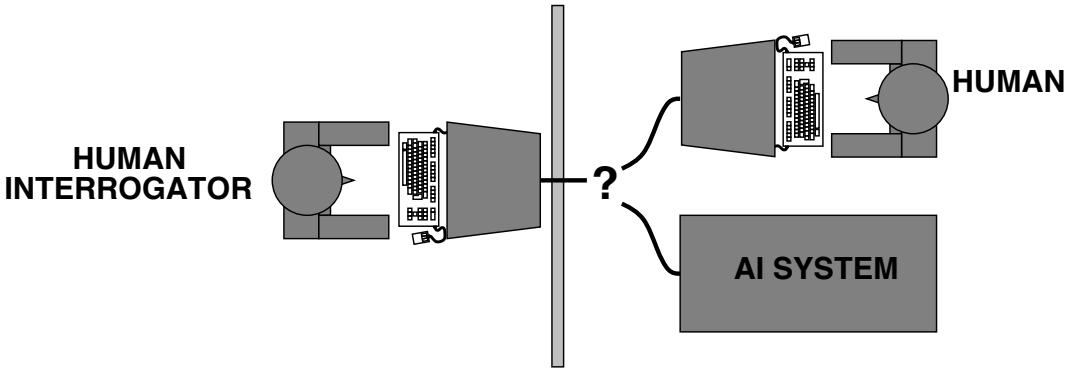
Direct line through mathematics and philosophy to modern AI

Problems:
1) Not all intelligent behavior is mediated by logical deliberation
2) What is the purpose of thinking? What thoughts **should** I have out of all the thoughts (logical or otherwise) that I **could** have?

# Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

- "Can machines think?" $\longrightarrow$ "Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game




- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not **reproducible**, **constructive**, or amenable to **mathematical analysis**

# Acting rationally

Rational behavior: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

> **Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good**

# Rational agents

An agent is an entity that perceives and acts

This course is about designing rational agents (più o meno ...)

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat: **computational limitations make perfect rationality unachievable**

$\rightarrow$ design best program for given machine resources

# Rational agents

Cosa significa razionale ?

- provvisito di ragione come facoltà peculiare dell'uomo (Devoto - Oli)

Ragione: facoltà di pensare, peculiare dell'uomo, soprattutto in quanto capacità di discernere, di determinare rapporti logici e di formulare giudizi (Devoto - Oli).

.... non proprio chiario cosa significhi in un contesto artificiale (senza assumere imitazione del comportamento umano ...)

# AI prehistory

| | |
|---|---|
| Philosophy | logic, methods of reasoning<br>mind as physical system<br>foundations of learning, language, rationality |
| Mathematics | formal representation and proof<br>algorithms, computation, (un)decidability<br>(in)tractability, probability |
| Psychology | adaptation<br>phenomena of perception and motor control<br>experimental techniques (psychophysics, etc.) |
| Economics | formal theory of rational decisions |
| Linguistics | knowledge representation<br>grammar |
| Neuroscience | plastic physical substrate for mental activity |
| Control theory | homeostatic systems, stability<br>simple optimal agent designs |

# Potted history of AI

| | |
|---|---|
| 1943 | McCulloch & Pitts: Boolean circuit model of brain |
| 1950 | Turing's "Computing Machinery and Intelligence" |
| 1952–69 | Look, Ma, no hands! |
| 1950s | Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist |
| | Proved most of theorems in Russell's Principia Mathematica |
| | Gelernter's Geometry Engine |
| 1956 | Dartmouth meeting: "Artificial Intelligence" adopted |
| 1965 | Robinson's complete algorithm for logical reasoning |
| 1966–74 | AI discovers computational complexity |
| | No domain information |
| | Failure of automatic translation: |
| | Lo spirito è forte ma la carne e' debole - = |
| | La vodka è buona ma la bistecca è marcia |
| | Computational complexity |
| | Models not adequate (e.g. perceptrons) |
| | Neural network research almost disappears |

# Potted history of AI

| | |
|---|---|
| 1969–79 | Early development of knowledge-based systems (DENDRAL, Domain Knowledge |
| 1980–88 | Expert systems industry booms |
| 1988–93 | Expert systems industry busts: "AI Winter" |
| 1985–95 | Neural networks return to popularity |
| 1988– | Resurgence of probability; general increase in technical deptl |
| | "Nouvelle AI": ALife, GAs, soft computing |
| 1995– | Agents, agents, everywhere . . . |
| 2003– | Human-level AI back on the agenda |

# State of the art

Which of the following can be done at present?

♢ Play a decent game of table tennis

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time
- ♢ Converse successfully with another person for an hour

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time
- ♢ Converse successfully with another person for an hour
- ♢ Perform a complex surgical operation

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time
- ♢ Converse successfully with another person for an hour
- ♢ Perform a complex surgical operation
- ♢ Unload any dishwasher and put everything away

# State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Telegraph Avenue
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at Berkeley Bowl
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time
- ♢ Converse successfully with another person for an hour
- ♢ Perform a complex surgical operation
- ♢ Unload any dishwasher and put everything away

# AI News

Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997

OTTER proved a mathematical conjecture (Robbins conjecture) unsolved for decades

No hands across America (driving autonomously 98% of the time from Pittsburgh to San Diego, 2797 miles)

During the 1991 Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people

NASA's on-board autonomous planning program controlled the scheduling of operations for a spacecraft

Proverb solves crossword puzzles better than most humans

# Unintentionally funny stories

One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe threatened to hit Irving if he didn't tell him where some honey was. The End.

Henry Squirrel was thirsty. He walked over to the river bank where his good friend Bill Bird was sitting. Henry slipped and fell in the river. Gravity drowned. The End.

Once upon a time there was a dishonest fox and a vain crow. One day the crow was sitting in his tree, holding a piece of cheese in his mouth. He noticed that he was holding the piece of cheese. He became hungry, and swallowed the cheese. The fox walked over to the crow. The End.

# Unintentionally funny stories

Joe Bear was hungry. He asked Irving Bird where some honey was. Irving refused to tell him, so Joe offered to bring him a worm if he'd tell him where some honey was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was . . .

# IA forte e IA debole

Ipotesi dell' **intelligenza artificiale forte**: è possibile costruire macchine che pensino realmente (e non solo simulino il pensiero) e facciano ciò che pensano.

– Solleva problemi concettuali e filosofici molto difficili

Ipotesi dell' **intelligenza artificiale debole**: è possibile costruire macchine macchine che agiscano come se fossero intelligenti (simulino il pensiero)

La maggior parte dei ricercatori in IA da per scontata l'ipotesi debole e non si preoccupa dell'altra

# La macchine possono pensare ?

Ovviamente dipende dalle definizioni .... ma la domanda è comunque mal posta

- le macchine possono volare ?
- le macchine possono nuotare ?

Le risposte dipendono dal linguaggio ....

(in russo le navi "nuotano")

# Il punto di vista di Turing

In "Computing machinery and intelligence" Turing osserva:

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 used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition **I shall replace the question by another**, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the **imitation game.**

# The opinion of Turing

So the question is:

**Are there imaginable digital computers which would do well in the imitation game**

the answer of Turing (in 1950) was yes:

**I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 10**9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning**. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

# Obiezioni possibili

Turing esaminò tutte le possibilie obiezioni alla sua visione (in pratica tutte quelle sollevate nei 50 anni successivi) ovvero:

- The Theological Objection
- The "Heads in the Sand" Objection
- The Mathematical Objection
- The Argument from Consciousness
- Arguments from Various Disabilities
- Lady Lovelace's Objection
- Argument from Continuity in the Nervous System
- The Argument from Informality of Behaviour
- The Argument from Extrasensory Perception

ne vediamo alcune

# Theological objection

Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think

It appears to me that **the argument quoted above implies a serious restriction of the omnipotence of the Almighty**. It is admitted that there are certain things that He cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this sort. An argument of exactly similar form may be made for the case of machines.

...

In the time of Galileo it was argued that the texts, "And the sun stood still . . . and hasted not to go down about a whole day" (Joshua x. 13) and "He laid the foundations of the earth, that it should not move at any time" (Psalm cv. 5) were an adequate refutation of the Copernican theory. With our present knowledge such an argument appears futile.

# "Heads in the Sand" Objection

The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so

I do not think that this argument is sufficiently substantial to require refutation. Consolation would be more appropriate: perhaps this should be sought in the transmigration of souls.

# Mathematical Objection

A number of results of mathematical logic can be used to show that there are limitations to the powers of discrete-state machines:

– Godel's incompletenes theorem (1931): in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent

– Turing (1937): there are certain things that such a machine (a digital computer) cannot do

If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply .... This is the mathematical result: it is argued that it proves a disability of machines to which the human intellect is not subject.

# Mathematical Objection (answer)

1) it has only been stated, without any sort of proof, that no such limitations apply to the human intellect. But I do not think this view can be dismissed quite so lightly.

Come confrontare un sistema formale con il pensiero umano ? Chi garantisce la consistenza del pensiero umano ? La storia è piena di errori (es. colorazione delle mappe, risultato del 1879 ... dimostrato sbagliato nel 1890 e dimostrato nel 1977)

2) This argument rest essentially on the condition that the machine must not make mistakes. But this is not a requirement for intelligence (see arithmetic progression and Gauss)

Lucas non può asserire in modo consistente che questa formula è vera

3) Il teorema di Godel in principio non si applica ai computer (che sono sistemi finiti)

# Arguments from Various Disabilities

I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X." Numerous features X are suggested in this connexion I offer a selection:

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new.

Oggi (50 anni dopo) vediamo che alcune cose sono state fatte

– scoperte in matematica, chimica ... – apprendimento dall'esperienza – uso e comprensione del linguaggio

Spesso si sbaglia a ipotizzare quali sono in processi mentali (intuizione, comprensione ... ) necessari per un determinato comportamento

# The Argument from Informality of Behaviour

It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances ... To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible.

There may however be a certain confusion between **"rules of conduct" and "laws of behaviour" ...** However, we cannot so easily convince ourselves of the absence of complete laws of behaviour as of complete rules of conduct. The only way we know of for finding such laws is scientific observation, and we certainly know of no circumstances under which we could say, "We have searched enough. There are no such laws."

Alla fine, noi tutti obbediamo a precise leggi fisiche (ad esempio quelle che regolano le interazioni fra atomi ...

# Lady Lovelace's Objection (1842)

Machine can never do anything really new: "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform"

Nihil sub solem novum ?!

Who can be certain that "original work" that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles.

**Learning**. The idea of a learning machine may appear paradoxical to some readers. How can the rules of operation of the machine change? They should describe completely how the machine will react whatever its history might be, whatever changes it might undergo. The rules are thus quite time-invariant. This is quite true. The explanation of the paradox is that the rules which get changed in the learning process are of a rather less pretentious kind, claiming only an ephemeral validity.

# The Argument from Consciousness (IA Forte)

This argument is very, well expressed in Professor Jefferson's Lister Oration for 1949, from which I quote. "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."

According to the most extreme form of this view the only way by which one could be sure that machine thinks is to be the machine and to feel oneself thinking ...

Likewise according to this view the only way to know that a man thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe "A thinks but B does not" whilst B believes "B thinks but A does not." Instead of arguing continually over this point **it is usual to have the polite convention that everyone thinks.**

# The Argument from Consciousness

Un dialogo classico:

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

**Witness**: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

**Witness**: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

**Witness**: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.

**Witness**: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

What would Professor Jefferson say if the sonnet-writing machine was able to answer like this ?

# IA forte

Sintesi dell'urea nel 1848: sostanze inorganiche trasformate in sostanze organiche

L'urea sintetizzata **è** urea

Il pensiero artificiale **è** pensiero ?

Searl (1980): nessuno pensa che la simulazione al computer di una tempesta ci debba bagnare. Perche' mai uno sano di mente dovrebbe supporre che una simulazione al computer di processi mentali debba realmente possedere dei processi mentali. La stanza cinese ...

La (simulazione della) addizione fatta dalla CPU è un'addizione ? Molti direbbero si ....

La CPU **sa** fare le addizioni ?

La risposta di Turing è l'educata convenzione: se si esibisce un comportamento pensante ammettiamo che ci sia del pensiero ... (anche fra gli umani)

# Un problema filosofico: mente-corpo

La mente e il corpo sono separati (dualismo) oppure no (monismo o fisicalismo).

Se si, come fa la mente a controllare il corpo.

Se no, stati mentali = stati fisici, come e' possibile ?

Decartes (1641): mente = processo privo di estensione spaziale o proprieta' materiali, separata dal corpo (interazione tramite ghiandola pineale)

# Fisicalismo, un problema

Stato mentale = configurazione fisica (ad esempio degli atomi nel cervell

So che sto mangiando una pizza = particolare configurazione

**Esperimento del cervello nella vasca** – cervello tolto alla nascita e fatto vivere in un ambiente artificiale, simulando i segnali dall'esterno e usando i segnali motori del cervello per modificare la simulazione – si potrebbe avere uno stato mentale "sto mangiando una pizza" ma questo non e' possibile

Contenuto allargato o ristretto ?

# Funzionalismo e naturalismo bilologio

## Funzionalismo

– Stato mentale = ruolo, cio funzione, relazione causale, rispetto ad altri stati mentali, percezioni e comportamento.

– Dunque gli stati mentali sono molteplicemente realizzabili, anche in modo artificiale

## Naturalismo biologico

– Stato mentali = caratteristiche di alto livello causate da processi fisici di basso livello nei neuroni

– Quello che conta sono alcune proprietà non fiunzionali dei neuroni

# Esperimento della protesi cerebrale

## Sostituisco uno a uno i neuroni di un cervello con dispositivi isomorfi

## Alla fine la coscienza rimane o no ?

Funzionalisti: Si (Moravec)

Naturalismo biologico: No (Searle: l'esperienza cosciente si riduce lentamente anche se il comportmento esterno rimane immutato).

# Funzionalismo

Per i funzionalisti nell'esperimento la coscienza rimane. IA forte possibile

Se la coscienza se ne andasse e il comportamento funzionale rimanesse:

i) eliminazione istantanea della volonta'

ii) dovremmo spiegare le manifestazioni di coscienza prodotte dal cervello artificiale facendo riferimento solo agli aspetti funzionali dei neuroni – questo dovrebbe applicarsi anche al cervello reale, che ha le stesse proprieta'.

Se pungiamo la persona avremo la risposta "Ahi! " sia con il cervello artificiale che con quello naturale

# Funzionalismo cont.

Due possibilità

1. i meccanismi causali della coscienza sono ancora operanti nella versione elettronica, che è quindi conscia
2. gli eventi mentali consci del cervello non hanno un collegamento causale con il comportamento manifestato: coscienza = epifenomeno, esiste ma non proietta alcuna ombra sul mondo osservabile

Se 2), allora anche nel caso naturale la risposta "Ahi" non deriva dalla coscienza ma da un altro meccanismo inconscio

# Naturalismo biologico

## Esperimento della stanza cinese (Searle)

Supponiamo che un computer si comporti come se capisse il cinese, passando il Test di Turing

In una stanza una persona **Che non sa il cinese** esegue (a mano) il programma che "comprende" il cinese, comunicando attraverso dipsotivi di I/O analoghi a quelli del computer.

La persona esibisce il comportamento di chi sa il cinese, ma non lo sa

# Naturalismo biologico cont.

Dunque:

– produrre gli otuput giusti non basta per "comprendere" (essere una mente)

–IA forte non possibile

Obiezione: se chiediamo alla stanza, la risposta è "so il cinese e lo posso dimostrare". Questo dovrebbe bastare

Searle: no, non c'e' comprensione del cinese nella persona, nelle regole e nella carta, quindi non ci può essere comprensione nell'insieme. Una proprietà del tutto deve stare da qualche parte (ma allora H2O ....)

# Searle

1. I programmi per computer sono entità sintattiche
2. Le menti umane hanno contenuto mentale (semantica)
3. La sintassi da solo non è costitutiva né sufficiente per la semantica
4. I cervelli causano le menti

Dai primi tre assiomi segue che i programmi non causano le menti ...

Ma c'e' una intepretazione non chiara di sintassi e semantica .... i segnali elettrici del cervello non sono sintassi alla fine ?

# Searle cont.

Il punto è che l'assioma 3) nega il funzionalismo, e quindi la possibilità che i non cervelli siano menti.

Guarda la stanza cinese ... dov'è la mente ?

Ma ... guarda i neuroni, dov'è la mente ?

La questione rimane aperta !

# A conclusion by Turing

**We may hope that machines will eventually compete with men in all purely intellectual fields.** But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. **This process could follow the normal teaching of a child.** Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.