

Philosophy of Artificial Intelligence

laboratory Gerstner

Introduction



∴ Artificial Intelligence explores and investigates various kinds of intelligent behaviour by attempting to create it in the artificial.

- such as problem solving, learning, natural language understanding, deduction, etc.
- the research vehicle is (experimenting medium) is computer

∴ according to Marvin Minsky Artificial Intelligence is a science of how to persuade computers to exhibit such a type of behaviour that conventionally requires Human Intelligence

Cognitive Science



∴ Artificial Intelligence belongs to the family of cognitive science, a border science including following aspects:

- cognitive psychology,
- neurology,
- philosophy of mind
- ethology
- logic
- evolutionary computing
- sociology and other

∴ Artificial intelligence is sometimes called machine intelligence, computational psychology or automated reasoning.

What Belongs to Artificial Intelligence



- ∴ Automated Reasoning, Theorem Proving and Game Playing
- ∴ Expert and Knowledge-Based Systems
- ∴ Natural Language Processing and Semantic Modelling
- ∴ Robotics, Planning and Scheduling
- ∴ Recognition (machine vision, speech recognition)
- ∴ Machine Learning
- ∴ Evolutionary Techniques
- ∴ Neural Networks and Parallel Distributed Processing
- ∴ Distributed Artificial Intelligence and Multi-Agent Systems
- ∴ Philosophy Artificial Intelligence

Three Schools of Artificial Intelligence



∴ symbolic functionalism

- intelligence represented in symbols and mutual manipulations
- formal models of reasoning, knowledge based

∴ connectionism

- inspired by natural processes
- emergency of intelligent behaviour

∴ robotics functionalism

- computational implementation of behaviourism

Symbolic Functionalism



Good Old Fashioned Artificial Intelligence – GOFAI

∴ Functional Hypothesis

Intelligent behaviour of a system is achieved through interaction among several system's components of different functionality

∴ Physical Symbol Systems Hypothesis

Physical Symbol Systems is an inevitable and satisfactory mean for implementation of Intelligent behaviour

Physical Symbol Systems Hypothesis



∴ Physical Symbol Systems (PSS) is a general machine providing evolution of population of symbol structures in time

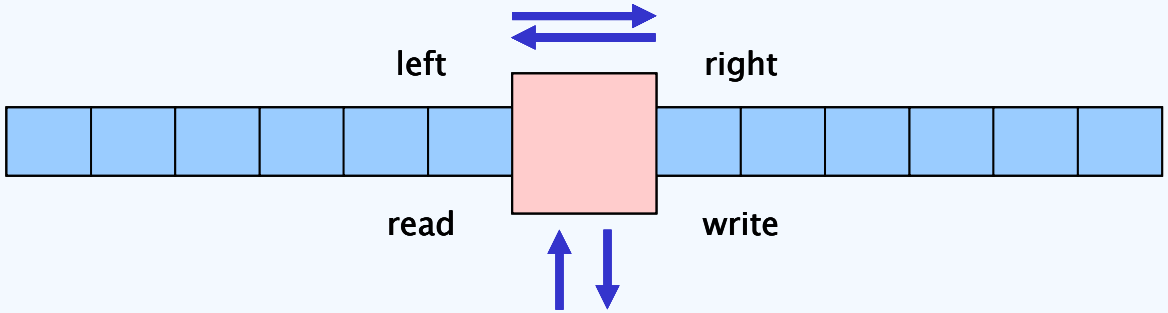
∴ Fundamental building primitives of the systems are symbols and operators of creation, destruction, modification and reproduction.

∴ an example of PSS is a Turing Machine

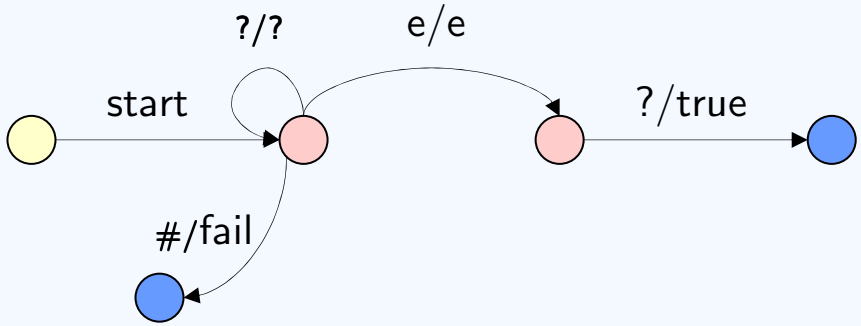
- with the tape as the only symbol structure and single operator of modification
- or with a symbol as a single symbol structure and applying operators of destruction and creation

∴ Symbolic Functionalism within AI concerns primarily methods for knowledge representation and reasoning models

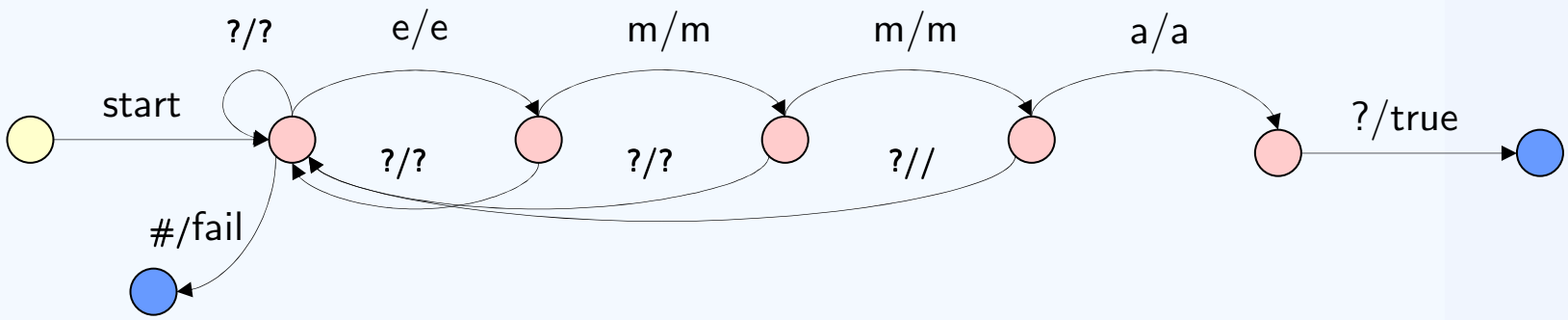
Turing Machine



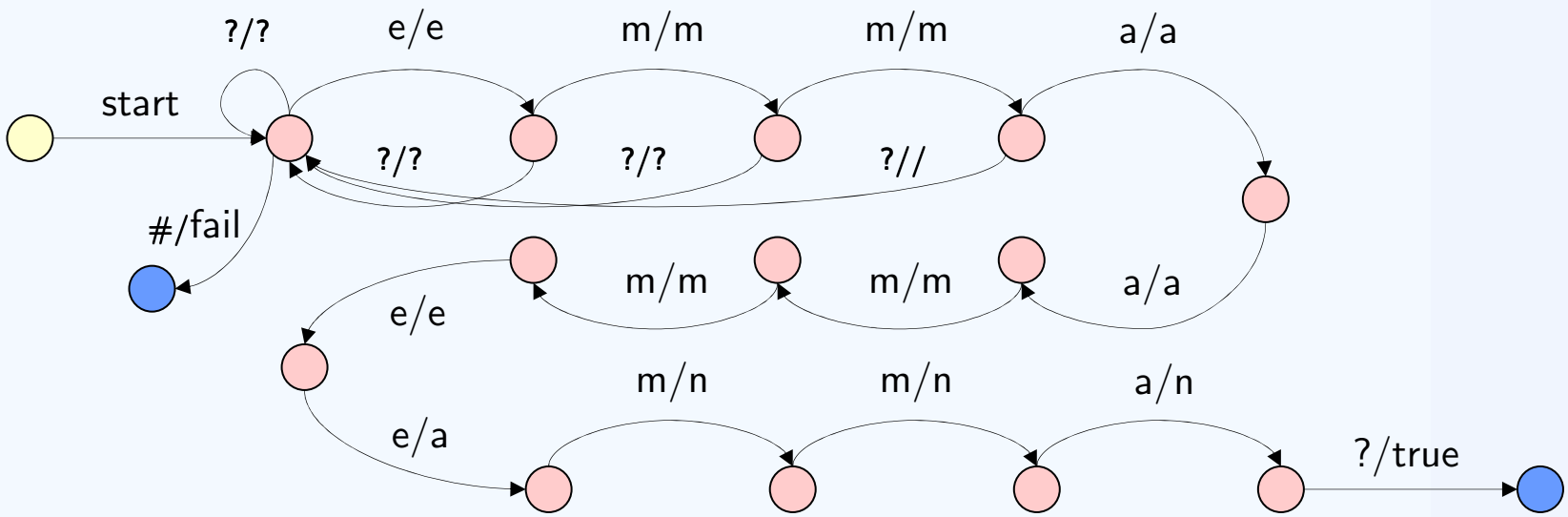
Turing Machine



Turing Machine



Turing Machine



Connectionism



∴ Essence of intelligence is rooted in static interconnection of huge number of simple processing units and in parallel behaviour of the system as a whole.

∴ Inspiration from the human brain operations that is a unarguably a medium of intelligent behaviour

∴ Neuron as a this simple processing unit gets excited according to weighted sum of incoming precepts

∴ Artificial Neural Networks belong to the field of connectionism

Connectionism cont'



∴ Isn't it just an instance of physical symbol system?

- intelligence is based on interaction among components
- components are same and they are static in time

∴ black box paradigm ~ we do not need to understand how the brain functions when solving a problem

- good for classification type of reasoning rather than deduction
- knowledge is not modelled

Robotic Functionalism



- ∴ It is based on computational implementation of the behaviourism as a psychological school of thought
- ∴ While symbolic functionalism investigates reasoning background behind intelligent behaviour robotic functionalism rather explores and simulates outer of intelligent behaviour
- ∴ Though the intelligence here is also rooted in interaction among systems components, unlike connectionism they are heterogeneous and hierarchically structured
- ∴ Intelligent behaviour is viewed as a sensible interaction among

agent × environment × task

spider paradigm

Emerging Fields in Artificial Intelligence



∴ Genetic Algorithms

- natural selection (crossover, mutation of population of solutions)
- due to its natural inspiration people classify genetic algorithms as a part of connectionism (it however does not seem to fit connectionism definition – sounds more like PSS)

∴ Distributed Artificial Intelligence

- concept of agency, collaboration between independent computational units
- resembles features of connectionism

∴ Artificial Life

- simulation of emergent 'life-like' properties in computer

Models of Understanding



let us view extent of artificial intelligence in terms of ability to understand

∴ we distinguish weak property of a system (its specificity) and strong property (its generality)

∴ weak understanding (Turing)

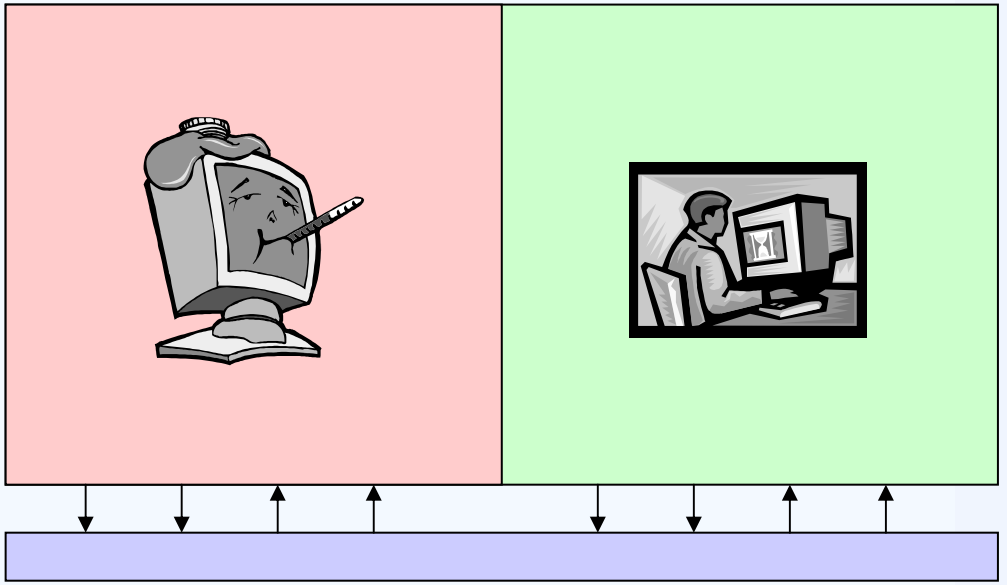
- such an understanding if inputs so that it makes the system to react intelligently (as human)

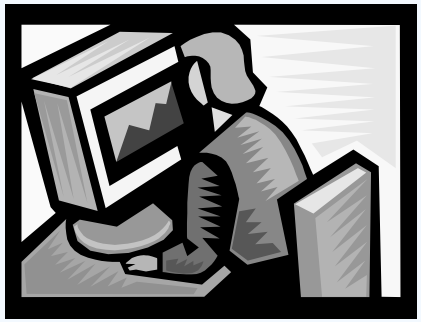
∴ strong understanding (Brentano)

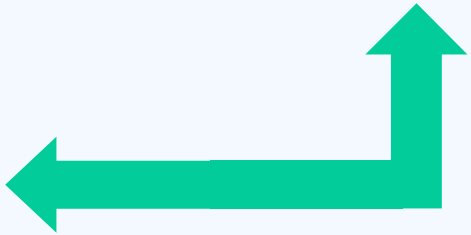
- such that the system feels the feeling of understanding as human do

∴ This is why we distinguish strong and weak artificial intelligence

Turing Test







Middling Artificial Intelligence



∴ According to Smith there is middling artificial intelligence, where the right type of behaviour reacting to given precepts is done through an appropriate knowledge structures and reasoning machinery.

- Robotics Functionalism – Weak Artificial Intelligence
- Symbolic Functionalism – Middling Artificial Intelligence
- Connectionism – Strong Artificial Intelligence (???)

∴ Example of natural language understanding

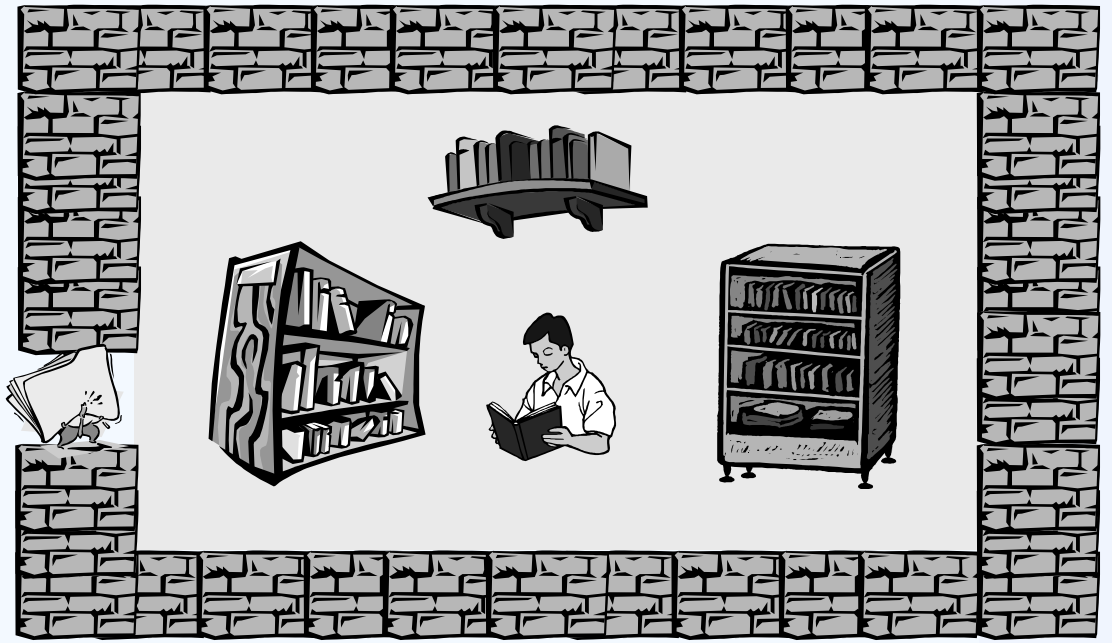
We claim that there is nothing like Strong Artificial Intelligence

Chinese Room Argument



Passing a Turing test does not imply strong artificial intelligence as there is nothing in the room understanding Chinese, though it looks like





Objections to Complete Models of Computational Reasoning



∴ Completeness Theorem (Godl): For any fact that (sentence in first order logic - FOL) logically follows from another fact, or a set of facts (theory), there exists a logical proof.

∴ However the entailment in FOL is semidecidable – we can prove that entailment follows from premises if it does but we cannot tell whether it does. We can end up in endless loop of constructing new and new theorems, while still not knowing whether it will lead anywhere.

∴ Halting Problem: Question hard to be decided by any program. Will execution of the program *P* eventually halt or will it run forever? Turing proved for any formal system *H* there exists such a program *P* that *H* will not be able to answer the halting problem correctly.

∴ Godl Incompleteness Theorem: Either man can solve more problems form the theory of numbers than a machine or there are problems that man cannot solve

Objections to Complete Models of Computational Reasoning



∴ From Godl incompleteness theorem follows: For any computer program there exists solvable problem that the program cannot solve.

∴ This does not prove that the man has intellectual superiority over machine

∴ Godl theorem refutes strong mechanistic hypothesis :

- There is a program that can answer all question that can be answered

∴ While it does not prove mentalist hypothesis (antithesis of above) :

- There is an answerable question that no program can answer

Approaches to Modeling Intelligent Reasoning



- ∴ theoretical reasoning – models of deduction, theorem proving ~ strong approach – high computational complexity (calculative rationality bottleneck)
- ∴ knowledge-base reasoning – (reactive) hardwired reasoning rules, sometimes with memory ~ weak approach (case specific)– reasonable complexity
- ∴ state space search – searching the space of possible solution with a goal to find a solution – high computational complexity – state space pruning by heuristics
- ∴ practical reasoning – reasoning in an intentional system – based on models of intentions, beliefs and desires – combination of knowledge-base reasoning (*deliberation*) and state-space search (*planning*)