PHIL256

# 1 Introduction

Cognitive science is the interdisciplinary adj. 各学科间的；跨学科的 study of the human mind and one of its most characteristic properties, intelligence. Intelligence, broadly speaking, is an ability that human beings exercise when they solve challenging problems.

Mainly concerned for intelligence:

1. To **identify the resources** that people deploy when they act intelligently
2. To **understand how the resources are deployed** in this process.

Classic Cognitive Science 认 知 科 学；理 性 科 学 : 1950-1980

1. People employ **symbolic representations** of the information in order to act intelligently
2. People deploy those symbols in thought by **processing those symbols.**

Modern Cognitive Science: after 1980s

-The view of intelligence has been challenged

-studies of the **human brain** have suggested: intelligence does not involve symbols in any significant ways.

- Some philosophers have challenged the claim that **symbol processing** is either necessary for intelligence or sufficient for it.

The aim of this course is to examine Cognitive Science as a way of accounting for human thinking and intelligence.

# 2 Introduction: The Cognitive Paradigm

* The **Central Thesis** of Cognitive Science is that thinking is like computation.

a digital computer computes by taking various **representations of information** and **performing calculations** with them.  The result might be some new information, such as the calculation of a final mark.

Thus, thinking, involved in intelligence, is **analogous** 类似的 to the computations performed by a digital computer. So, in this analogy brain = CPU

However, this analogy **does not imply** that your brain **physically resembles** a digital computer.

the Central Thesis is more like what the philosopher of science Thomas Kuhn called a **paradigm**范例.

**A scientific paradigm fulfills at least the following functions:**

1. it tells researchers what sorts of phenomena to investigate, and what sort to ignore.

2. it tells them what sorts of theories to test, and what sort not to test

3. it tells them how to perform the tests, and how not to.

Central Thesis acts as a paradigm for Cognitive Science:

1. it tells researchers to investigate **intelligent behaviors** (phenomena).

2. it tells them to theorize about **mental representations and procedures** (theories).

3. it tells them to test theories **using computational models, experiments** (How ), etc.

## 2.1 intelligence

What require intelligence?

1.puzzle solving,

2. **Argumentation**, that is, giving strong reasons for holding views, e.g., in science, mathematics, philosophy, etc.

3. **Technological development**, e.g., engineering, computer programming, etc.

In Classic Cognitive Science, **intelligent thinking is any thinking in which knowledge or expertise plays a major role**.

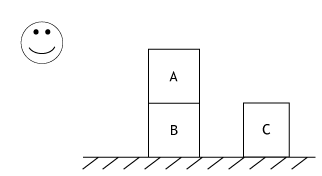
In order to play chess, you have to know the rules of the game and have some knowledge of strong patterns of play or "gambits", such as special openings.

On this view, **intelligence is knowledge-intensive**: you need lots of knowledge to be intelligent, and the more knowledge you have, the more intelligent your behaviour will be.

Thus, one of the main aims of Cognitive Science is to define what knowledge is and how it participates in thinking.

## 2.2 Mental Representations and Procedures

Have knowledge of something is to possess information about it in the form of **mental represenatations**.



The observer may form a **mental representation** of this scene in his mind. The mental representation might take the form of some **statements** about where the blocks are in relation to each other, e.g.,

1. Block A is on block B.
2. Block B is on the ground.
3. Block C is on the ground.
4. Block C is right of block B.

Mental representation of a scene, is A set of statements about the scene in the mind of the observer

**Mental procedures** are methods that our minds possess for manipulating mental representations of knowledge in order to achieve certain outcomes.

ex Suppose that the observer of the block scene above wants to spell the word "cab" with the blocks. To satisfy this goal, the observer might come up with a plan by applying the following rules:

To have block x on block y, place block x on top of block y.

To place block x on top of block y, remove other blocks from on top of y, pick up block x, move it on block y and let go of block x.

Using these rules, the observer could work out and execute a **plan** that would allow him to realize his goal.

\*\*the point is simply that we can think of mental representations and procedures as ways in which **knowledge is packaged and applied in our minds**.

## 2.3 Theory Assessment in Cognitive Science

In Cognitive Science, a theory is a **model or explanation** for how it is that people act intelligently in a given situation.

**a theory** in Cognitive Science makes claims about what **mental representations** are active in the situation and what **mental procedures** are used to apply the representations.

In the block scene example above, the theory outlined makes the claim that arranging the blocks in a certain way calls for a **set of statements** that constitute knowledge of the arrangement of blocks and a **set of rules** for constructing a plan in view of that knowledge.

write a computer program that contains **an analogous set of statements and rules** and see if it can reproduce the intelligent behaviour that humans display when confronted with this problem.

Run the program and see what it does. If its behaviour does match the behaviour of humans, then the theory is **confirmed** to some extent. If its behaviour is unlike the behaviour of humans, then the theory is **disconfirmed**.

This is known as the **cognitive modeling,** it is an approach to theory assessment.

Cognitive Science is a **highly interdisciplinary pursuit** including researchers from many academic disciplines such as Anthropology, Linguistics, Neuroscience, Philosophy, and Psychology.

## 2.4 summary

Central Thesis of Cognitive Science is that thinking is like computation on a digital computer.

**This thesis is CRUM** (Computational Representational Understanding of Mind)

As a paradigm, CRUM implies several assumptions about the nature of intelligence and how to explain it,

1. Intelligence is knowledge-intensive,
2. Intelligence is produced by mental representations and procedures, and
3. Theories about intelligence may be tested in various ways, especially by computational models.

 A **good record of success, good prospects for future success**, and **performance superior to other paradigms** are essential for the continued acceptance of CRUM as a picture of how human intelligence works.

## 3.1 Prehistory

The ancient Greeks, e.g., Plato and Aristotle, laid out the questions (and some of the sorts of answers) that cognitive scientists continue to pursue. The main questions are:

1. What do you know and how do you know it? (**epistemology**)
2. What kind of thing is a mind? (**metaphysics**)
3. How does a mind give rise to thinking? (**psychology**)

**Plato** (ca. 400 BC) gave the following answers (roughly speaking):

1. Knowledge is the grasp of true ideas (extrasomatic) and is obtained by the mind in a disembodied state.
2. A mind is collection of physical (**hydraulic**) and non-physical (immortal) organs.
3. Physical thinking is the interaction among mental organs via pressures and ratios of organic fluids.

**Locke** presents a very different model of human cognition (ca. 1690):

1. Knowledge is a set of general and tentative propositions gathered from sensory experience.
2. A mind is like a sheet of white paper onto which ideas are written by experience.
3. Thinking is a process of reflection in which ideas are abstracted from sensations.

In the 1940s, Wiener and others introduced a new model of mind called **cybernetics** based on an analogy with automatic range finders for anti-aircraft guns:

1. Knowledge is the information used to make predictions about futures states of the environment.
2. A mind is a set of feedback mechanisms connecting a body's sensory and motor apparatus.
3. Thinking is the adaptation of future conduct based on evaluation of past performance.

The **control system analogy** declined with the advent of cognitive science, but is now returning in a more advanced form in dynamical systems theory and robotics, which we will discuss later on.

# Module 2 Computing Machinery and Intelligence (Turing)

**CRUM is not so much a theory about human thinking and intelligence as a paradigm. That is, it provides a framework for identifying important questions to answer, for composing theories to answer those questions, and for how those theories should be assessed.**

**Alan Turing's paper** Computing machinery and intelligence(1950).

In the 1930s and 40s, Turing was the prime developer of the **theory of computation** and therefore one of the "founding fathers" of computer science.

Turing's paper was one of the earliest papers to **lay out the rationale of this program and to explore its assumptions and the general objections to such a program of research**.

## 2.2 Alan Turing

Alan Turing (1912-1954) was a British mathematician, noted for his development of modern concepts of computation.

In World War II, he worked secretly at Bletchley Park and was there instrumental in breaking the code of the Nazi Enigma machine, which the Nazis used to send secret messages to military leaders in the field and to spies abroad.

After the war, he moved on to Manchester University, where he helped to design the Mark I, one of earliest, general-purpose digital computers. By 1950, he was no longer very interested in the design of computers and much more interested in their potential mental abilities.

## 2.3 Computing Machinery and Intelligence

Turing proposes to address the question, "**can machines think?**"

Turing proposes a different method for addressing the question, namely the **imitation game**.

The players of the game are a man (A), a woman (B), and an interrogator (C) who may be either male or female.

The setup goes as follows: The interrogator sits in a different room apart from A and B. The interrogator can put written questions to A and B, whom he addresses anonymously as X or Y, with the object of determining which one is the man and which the woman, based presumably on the answers.

So, if the interrogator is able to distinguish reliably between male and female players, then it will be because he or she is able to detect some **interesting and important aspects** of gender differences.

**Turing Test**

Turing suggests an analogy with the question of whether machines can think. Suppose that you substitute a computer for player A and **change the goal** of the interrogator to the determination of which player is the computer and which the human being. If the interrogator can reliably tell the difference between the computer and the human being, then you would conclude similarly that he or she is able to detect some interesting and important aspects of the difference between humans and machines. One obvious candidate for this difference would be that the human being can think intelligently whereas the computer cannot. By the same token, if the interrogator cannot distinguish reliably between the computer and the human being, then that result would suggest that **there is no interesting and important feature that distinguishes human from computer**, where their ability to conduct a conversation is concerned. In that case, Turing suggests, it is reasonable to conclude that the computer can think.

## 2.4 Critique of the New Problem

Imitation game is suitable for the purpose of determining whether or not a computer can think as a person.

To ask whether or not the computer has a brain or whether or not the computer looks like a human being would amount to begging the question. Such a procedure would assume that thinking requires a brain or requires the thinker to be a member of a particular biological species. The problem we face is whether or not such assumptions are, in fact, true.

This point indicates ways in which we should **not judge the issue**.

Turing argues that such a test is one that allows the computer to overcome its superficial differences to a human being. In other words, **it can disguise itself as a human** when communication with it is limited to text messages.

Suppose C asks the computer to add 34957 to 70764. The computer might have the sum much faster than a real human would have it, but the computer could pause for a little while in order to disguise this fact.

**Conversation is an activity that normally allows us to gauge the intelligence and thinking capacity of other people**. After conversing with someone for a while, you may have a fair impression of how smart that person is.

As a result, we would want to draw a conclusion only from a fair amount of experience with the imitation game instead of from a single trial.

## 2.5 digital computer

A typical computer contains the following three components:

**Memory** (store):  place in which the computer can keep information as required for it to perform its calculations.

**Executive unit**: a component that coordinates the activity of the rest of the computer. In a typical modern computer, the executive unit is the Central Processing Unit (CPU),

**Control**: a set of instructions that tells the executive unit what calculations to carry out, and in what order.

The instructions that comprise a computer program are not like the instructions that you would find in English in a recipe book, say. Instead, a computer program is simply a sequence of numbers that the executive unit is able to interpret as instructions for manipulating the computer's components.

It is by **carrying out a sequence** (possibly very, very large) **of such instructions that the computer's behaviour is determined**.

## 2.6 Universality of Digital Computers

a **general-purpose digital computer** is one that can imitate the activity of any other digital computer.

Put another way, a general-purpose digital computer is capable of running any program that can be run on another digital computer.

 In that case, we could conclude that the intelligence of the computer, **its ability to think, is a fact about the program and not a fact about the physical machine** involved.

That conclusion would be very interesting indeed because it would suggest that intelligence is a highly abstract thing that all sorts of physically different beings might possess, as opposed to a highly concrete thing that only, perhaps, human beings happen to possess.

## 2.7 Contrary Views on the Main Question

We can now consider some objections to this prospect as discussed by Turing.

The **first objection is the theological objection**.

* Argument: thinking requires a soul, computers have no souls, so computers cannot think
* Reply: it only follows that computers **do not** think
  + Being omnipotent adj. 无所不能的, God **could** give souls to computers, enabling them to think
* Ultimately, whether computers can think is an empirical adj. 经验主义的，完全根据经验的 matter, determined by empirical tests (e.g., the imitation game)
  + Biblical arguments about empirical matters are unreliable

Turing replies that, among other things, this conclusion does not follow from the premises. The premises state merely that people have souls, which allow them to think, whereas computers do not have souls. What follows is that computers do not think, not that computers cannot think. To say that computers cannot think, Turing argues, is to say that God could not give a computer a soul. Yet, if God is omnipotent, then why could He not give a computer a soul?

The **second objection is the mathematical objection**

**The mathematical objection** relies on some results of the theory of logic and computation.

The idea behind this objection is that here is a question that a human can answer but that a given computer cannot.

* Some questions are answerable by humans and not by digital computers:
  + Gödel's theorem shows that there are questions of logic not answerable, in principle, by a given computer
  + The person framing the question can determine the answer
* Reply: there may be such questions for any given human
  + Perhaps a computer could scan your brain and frame a question unanswerable by you, in principle
  + The computer could compute the answer though
* The Gödel argument begs the question

Gödel's proof does not demonstrate anything about what a given human being can or cannot know. As a result, the objection **begs the question**: we cannot claim to know that we are in a superior position to digital computers in this matter.

* We can know that something thinks only if we know that it has conscious adj. 意识到的；故意的；神志清醒的 experiences
  + It is like something to be intelligent
* Reply: We know of **our own** conscious experiences only
  + Solipsism n. 唯我论: only I am known to be conscious
* To avoid absurdity n. 荒谬；谬论；荒谬的言行, we must admit behavioural evidence for intelligence
  + E.g., the imitation game

The **third objection is the consciousness objection**.

**The argument from consciousness** holds that we cannot say that something thinks unless we know that it has conscious experiences.

The consciousness objection implies that thinking requires phenomenological experience to occur. In brief, thinking intelligently should be accompanied by a special experience, perhaps a form of awareness of one's own thoughts.

The consciousness objection implies that thinking requires phenomenological experience to occur. In brief, thinking intelligently should be accompanied by a special experience, perhaps a form of awareness of one's own thoughts.

The **fourth objection is from Lady Lovelace**.

**Lady Lovelace** was a patron of Charles Babbage, who worked on building a mechanical computer called the Analytical Engine in the mid 19th century. Lady Lovelace noted that the Engine had no pretension to originality; it would do only whatever Babbage programmed it to do. Of course, as Turing points out, it may simply not have had the capacity to do anything original.

The **fourth objection is from Lady Lovelace**.

**Lady Lovelace** was a patron of Charles Babbage, who worked on building a mechanical computer called the Analytical Engine in the mid 19th century. Lady Lovelace noted that the Engine had no pretension to originality; it would do only whatever Babbage programmed it to do. Of course, as Turing points out, it may simply not have had the capacity to do anything original.

**Turing has two responses**.

1. This objection **begs** 回避正题 **the question**: How do we know that people ever do anything original? For all that we know, the things that we produce in our lifetimes may simply be the perfectly predictable in advance if we had all of the relevant information. If original means unpredictable in principle, then we cannot be sure that we are any different from computers in that respect.
2. If by original we simply mean surprising, then the objection is simply false. As Turing says, "machines take me by surprise with great frequency". One of the main reasons why people program computers is because they do not or cannot figure out the result of running the program for themselves. As a consequence, the results are sometimes surprising. A skeptic might then argue that the result is at least **predictable in principle**, so that the surprise is an accidental instead of a necessary feature of running the program. Here we have returned to the question-begging objection above: How do we know whether or not the surprises that people themselves produce are predictable in principle?

**To conclude**, none of this discussion should be taken to minimize the problem of whether or not machines can think. Even if we accept Turing's arguments and his rebuttals to objections, the issue remains decidable only through a rigorous empirical research program.