

The image shows a spiral-bound notebook with a light brown, textured cover. The spiral binding is on the left side. The text is centered on the cover.

# Artificial Intelligence

## Philosophy I

### Lecture 12

(24 November, 1999)

Tralvex (Rex) Yeap MAAI MSCS  
University of Leeds

# Content: Philosophy I



## 📄 Quick Review of Lecture 11

## 📄 Why Philosophy is a Weak Topic Among Previous/Earlier Batches?

## 📄 Introduction to Philosophical Aspects of AI

## 📄 List of Papers for Philosophy I

1. Computing Machinery and Intelligence
2. Glossary of Philosophical Terms
3. Mapping Great Debates: Can Computers Think?
4. The Turing Test & Chinese Room Experiment
5. Misguided Artificial Intelligence: The Turing Test
6. Twelve Reasons to Toss the Turing Test

## 📄 List of Papers for Philosophy I (cont)

7. Philosophy of AI: Part of Contemporary Philosophy of Mind - An Annotated Bibliography
8. How to Pass the Turing Test by Cheating
9. Lessons from a Restricted Turing Test
10. Chatterbots, Tinymuds, And The Turing Test: Entering the Loebner Prize Competition
11. Introducing MegaHal
12. HeX Template
13. Joseph Weintraub & His Therapist
14. Chess is Too Easy

## 📄 What's in Store for Lecture 13

# Quick Review on Lecture 11



- 📄 Overall Picture of Linguistics and Language Processing
- 📄 NLP Stages in Producing an Internal Representation of a Sentence
- 📄 Context Free Grammars
  - Components of a Natural Language
  - SMC Example: Sue, mouse & the cat
  - Top-down parsing (A)
  - Bottom-up parsing (A)
  - Top-down parsing (B) - using parse tree
  - Bottom-up parsing (B) - using parse tree
  - JA Example: John and apple
  - Alternative CFG Visualization
  - Proving that a list of words is a sentence
- 📄 Class Workout 1, 2 & 3: AI Exam 1995, 1996 & 1997 NLP questions
- 📄 Chomsky's Grammar Hierarchy
  - Type 0, 1, 2 & 3
- 📄 Grammar for Natural Language
  - Attribute-Value Grammar
- 📄 Semantics Interpretation
  - Logical Form
  - Logical Form of Phrases Examples
  - Lambda Calculus
  - Reduction Rules
  - $\alpha$ -reduction
  - $\beta$ -reduction
  - Semantics Annotations
  - Semantics Ambiguity
- 📄 Conclusion
- 📄 Class Activity 1, 2 & 3

# Why Philosophy is a Weak Topic Among Previous/Earlier Batches?



- Materials used by earlier classes (by another lecturer) were solely based on the **two classical papers** (Alan Turing's and John Searle's) - these are grossly insufficient.
- Past batches may not be used to reading research papers, and AT's and JS's paper are typical research papers.
- It was unfortunate that materials on the formal instantiation of Turing's test - **Loebner Prize Competition was overlooked.**
- A glimpse on the past AI exam papers would review that **newer philosophical topics** were covered, including discussion revolving the chess match by Deep Blue and Gary Kasparov.

# Introduction to Philosophical Aspects of AI



- ☞ **How is it possible for a physical thing--a person, an animal, a robot--to extract knowledge of the world from perception and then exploit that knowledge in the guidance of successful action?** That is a question with which philosophers have grappled for generations.
- ☞ It could also be taken to be **one of the defining questions** of Artificial Intelligence.
- ☞ **AI is, in large measure, philosophy.** It is often directly concerned with instantly recognisable philosophical questions: What is mind? What is meaning? What is reasoning, and rationality? What are the necessary conditions for the recognition of objects in perception? How are decisions made and justified?

## Introduction to Philosophical Aspects of AI (cont)



- ☞ Philosophy has a **more direct relation** to AI than it has to other sciences.
- ☞ Both subjects require the **formalization** of common sense knowledge.
- ☞ We may regard the subject of artificial intelligence as **beginning with Turing's paper** Computing Machinery and Intelligence (1950) and with **Shannon's** (1950) **discussion** of how a machine might be programmed to play chess.

## Introduction to Philosophical Aspects of AI (cont)



- ☞ In fact, much AI already builds on work by philosophers. An obvious example is the use of **speech act theory**, developed originally by philosophers such as John Austin, John Searle and Paul Grice.
- ☞ There are also various uses of **specialised logics**, e.g. deontic logic, epistemic logic, and modal logics, originally developed by philosophers in an attempt to clarify concepts like 'permission' and 'obligation' (deontic logic), 'knows' and 'believes' (epistemic logic), and 'necessarily' and 'possibly' (modal logic).
- ☞ These contributions from philosophy are not passively accepted in AI: putting them to use in designing working systems often **reveals shortcomings** and **suggests further development**.



## Introduction to Philosophical Aspects of AI (cont)



- ❏ Older contributions from philosophy includes Kant's proof in Critique of Pure Reason that learning from experience was impossible without some sort of prior (innate) conceptual apparatus.
- ❏ Another was Frege's heroic (but unsuccessful) attempt a century ago to show that all arithmetical concepts could be reduced to logical concepts and all arithmetical knowledge could be derived from logical axioms and rules. This led him to a number of extremely important results, including the first ever accurate analysis of the role of variables in mathematical expressions, discovery of the notion of higher order functions and invention of **predicate calculus**.



## Introduction to Philosophical Aspects of AI (cont)



- ☞ This led (via work by Russell, Church and others) to **lambda calculus**, **type theory**, and other important notions in computer science and formalisms for AI.
- ☞ More recently the old philosophical controversy about varieties of **forms of representations** which has become a topic of active AI research.
- ☞ Another recent development is recognition of deep connections between the AI task of understanding what sort of knowledge an intelligent system requires and the older philosophical activities of metaphysics, especially what Strawson described as 'descriptive metaphysics', including **ontology**, the attempt to characterise in a systematic way what exists.

# List of Papers for Philosophy I



1. Computing Machinery and Intelligence
2. Glossary of Philosophical Terms
3. Mapping Great Debates: Can Computers Think?
4. The Turing Test & Chinese Room Experiment
5. Misguided Artificial Intelligence: The Turing Test
6. Twelve Reasons to Toss the Turing Test
7. Philosophy of AI: Part of Contemporary Philosophy of Mind - An Annotated Bibliography
8. How to Pass the Turing Test by Cheating
9. Lessons from a Restricted Turing Test
10. Chatterbots, Tonymuds, And The Turing Test: Entering the Loebner Prize Competition
11. Introducing MegaHal
12. HeX Template
13. Joseph Weintraub & His Therapist
14. Chess is Too Easy

# Paper 1: Computing Machinery and Intelligence



☞ The Imitation Game

☞ Critique of the New Problem

☞ The Machine Concerned in the Game

☞ Contrary Views on the Main Question

1. The Theological Objection (God)
2. The 'Heads in the Sand' Objection (Superiority of Man)
3. The Mathematical Objection\* (Godel Theorem)
4. The Argument from Consciousness\* (Concerto)
5. Arguments from Various Disabilities (can't do X)
6. Lady Lovelace's Objections\* (No Originality)
7. Argument from Continuity in the Nervous System (Discrete State)
8. The Argument from Informality of Behaviour (No Exhaustive Rules)
9. The Argument from Extrasensory Perception (ESP)

*\* Appeared in 1995 paper*

# Paper 1: Computing Machinery and Intelligence (cont)



## The Mathematical Objection

**Summary** There are well known limits to the power of computers. For example, the halting problem (it is not possible to write a program that determines, in general, whether another program will ever terminate), so there will be some questions to which the machine will give a wrong answer, or never answer.

**Turing's reply** (1) We too give wrong answers. (2) It has not been shown that there is no halting problem for humans.

## The Argument from Consciousness

**Summary** Computers can never feel or have emotions.

**Turing's reply** The machine can reply to questions in a way that would suggest it has emotions.

# Paper 1: Computing Machinery and Intelligence (cont)



## Lady Lovelace's Objection

**Summary** Lady Lovelace, when writing about Babbage's Analytical Engine, claimed that machines could never do anything new, they could only do what they had been programmed for, and if we know the program then we will always know how the machine will behave. Turing paraphrases this by "A machine cannot take us by surprise."

**Turing's reply** You can still experience surprise in the behaviour of a formal system, even when you know all of its basic parts. To know a fact does not mean that you know all the consequences of that fact.

## The Argument from Informality of Behaviour

**Summary** (1) It is not possible to formulate rules that govern all aspects of human behaviour. (2) A machine will follow fixed rules whose nature can be determined by observing its behaviour.

**Turing's reply** (1) Because we cannot formulate rules that govern all of human behaviour does not mean that there are no such rules. (2) It is possible to write quite simple programs which a human observer would not be able to reconstruct on the basis of observation of external behaviour.

# Paper 1: Computing Machinery and Intelligence (cont)



## Other Criticisms

We might argue that the simulation of an intelligent, self-aware, thinking being does not necessarily mean that intelligent, self-aware thinking and understanding is really taking place.<sup>2</sup>

Turing would argue that this should not concern us, as from external observation alone you would not be able to prove that the machine was not self-aware, thinking, and understanding.

## Learning Machines

Turing suggests that one way such a 'thinking' machine may be produced is not to seek to emulate an adult mind, but rather to simulate a child's. That is, we could seek to create a machine that is capable of learning through education and experience.



## Paper 2: Glossary of Philosophical Terms



📄 A good A-Z list of philosophical terms which one would encounter during the study / reading of Alan Turing, John Searle paper and other handouts.

### ***a priori-a posteriori distinctions***

an epistemological distinction. *A priori* propositions, unlike *a posteriori* propositions, do not require to

### **abstract ideas**

Locke's attempt, rejected by Berkeley, to explain how an idea can stand for individuals of a given kind

### **abstract objects**

objects, such as numbers or universals, that do not exist as spatio-temporal particulars. Philosophers di

### **ad hominem argument**

a fallacious argument attacking the holder of a view rather than the position itself *or* a sound argument

### **alientation/estrangement**

Hegelian concept, also used by Marx and later European philosophers, to stand for a state of being cut

### **altruism**



# Paper 3: Mapping Great Debates: Can Computers Think?



- ☞ A set of 7 poster- sized argumentation maps that chart the entire history of the debate.
  - ☞ The maps outline arguments put forth since 1950 by more than 380 cognitive scientists, philosophers, artificial intelligence researchers, mathematicians, and others.
  - ☞ Every map presents 100 or more major claims, each of which is summarized succinctly and placed in visual relationship to the other arguments that it supports or disputes. The maps, thus, both show the intellectual history of this interdisciplinary debate and display its current status.
- ☞ Claims are further organized into more than 70 issue areas, or major branches of the arguments.
  - ☞ Here is a list of the seven maps:
    - Map 1** : Can computers think?
    - Map 2** : Can the Turing test determine whether computers can think?
    - Map 3** : Can physical symbol systems think?
    - Map 4** : Can Chinese Rooms think?
    - Map 5** : Part 1: Can connectionist networks think?
    - Map 5** : Part 2: Can computers think in images?
    - Map 6** : Do computers have to be conscious to think?
    - Map 7** : Are thinking computers mathematically possible?

# Paper 4: The Turing Test & Chinese Room Experiment



## The Turing Test

- Descartes' Challenge
- The Imitation Game - Judge, Man, Woman
- Turing's Test - Judge, Computer, Person
- Modified Turing Test Setup - Judge and Computer/Human
- Partial Turing Tests - Judge, Specific Function/Calculation

## Turing First Predictions - 50 years achievement

## Turing Second Predictions - Conversational Systems that does not contradicts.

# Paper 5: Misguided Artificial Intelligence: The Turing Test



 Introduction

 The Tricks

 The Chinese Room

 A Misdirection

 Conclusion

## Paper 6: Twelve Reasons to Toss the Turing Test



- ❏ Bad definition
- ❏ Credits to human not machine iff it works
- ❏ Does nothing to guide AI
- ❏ Existence of better definitions of intelligence
- ❏ Persistency of unbelievers
- ❏ A statistical misunderstanding
- ❏ Fatally subjective
- ❏ Should built on the theory of mind
- ❏ Easy to fool
- ❏ Least use of any theory on what it takes to build an intelligent machine
- ❏ Only algorithmic structure produces intelligence?
- ❏ Language bias

## Paper 7: Philosophy of AI: Part of Contemporary Philosophy of Mind - An Annotated Bibliography



 A good annotated compilation on bibliographies related to:

- (1) The Turing Test - 19 papers
- (2) The Chinese Room - 50 papers
- (3) Philosophy of AI, Misc - 25 papers
- (4) and other papers related to philosophy+AI specific topics

## Paper 8: How to Pass the Turing Test by Cheating



An interesting paper on how to pass the Turing Test by brute force without using any AI methods.

- 📄 Introduction
- 📄 Can Machines Think?
- 📄 Introduction
- 📄 Can Machines Think?
- 📄 The \$100,000 Question
- 📄 Early Optimism: ELIZA, PARRY & Shrdlu, PARRY Visits the DOCTOR
- 📄 Not Much has Changed: PC Therapist, TIPS & FRED

- 📄 How to Spout Gibberish
- 📄 How to Fool the Proles: Analysis, Design, Implementation
- 📄 Why the Loebner Contest is Doomed to Failure
- 📄 Conclusion
- 📄 References

## Paper 9: Lessons from a Restricted Turing Test



- 📄 The Turing Test and the Loebner Prize
- 📄 The Event
- 📄 Analysis
- 📄 Why a Loebner Prize?
- 📄 Alternatives to the Loebner Prize
- 📄 Postscript



# Paper 10: Chatterbots, Tinymuds, And The Turing Test: Entering the Loebner Prize Competition



- 📄 Introduction
- 📄 History
- 📄 Tinymud
- 📄 The Loebner Prize
- 📄 The CHATTERBOT Design
- 📄 Strategies for Conversational Programs: ‘Tricks’
- 📄 Does The Loebner Prize Further AI?
- 📄 Conclusion

# Paper 11: Introducing MegaHal

---



- Introduction
- The Loebner Contest
- Conversational Simulators
- MegaHal
- Conclusion

## Paper 12: HeX Template

---



- 📄 HeX was written purely to win the 1996 Loebner Contest, and it was successful.
- 📄 It is merely an Eliza clone, but the abusive personality.
- 📄 This is the actual template for all conversations.

## Paper 13: Joseph Weintraub & His Therapist



- History of the PC Therapist
- The Loebner Prize
- The PC Therapist III: Technology
- The PC Therapist IV
- Ordering Information

# Paper 14: Chess is Too Easy



Introduction

Creativity Ex Machina?

Deep Story

Forever Conscious

# What's in Store for Lecture 13



## Philosophy II



## Strong AI vs Weak AI

## List of Papers for Philosophy I

1. Minds, Brains, and Programs (John Searle, 1980)
2. The Chinese Room Argument (The Internet Encyclopedia of Philosophy)
3. Minds, Brains and Programs (Intro to Philosophy)
4. Searle's "Chinese Room" Argument (Indiana, 1990)
5. Minds and Machines (Monash)
6. John R. Searle's Chinese Room (Helsinki)

A spiral-bound notebook with a light brown, textured cover and a dark brown border. The spiral binding is on the left side. The text is centered on the page.

**End of Lecture 12**

Good Night.