

University of California, Merced
COGS 125 / CSE 175 : Introduction to Artificial Intelligence

Philosophy Of Artificial Intelligence

David C. Noelle, Ph.D.

September 10, 2019

Outline

- Reading Quiz
- Class Business
- Residual Questions
- A Brief Overview of the Philosophy of AI

Reading Quiz

- Readings:
 - any previously assigned readings
 - AlMA, Chapter 26
 - “March of the machines”, *The Economist*, June 25, 2016

Class Business

- Please visit the course web site on CatCourses regularly. Lecture slides are available there.
- Laboratory sessions were held for the first time last week. This week's sessions will review Java™ programming concepts. Attendance at laboratory sessions is required.
- Please resolve issues concerning appropriate preparation for the class as soon as you can.
- Please keep up with the readings, as specified in the course syllabus.

Course Logistics



History of AI & Math Review



Philosophy of AI

- Prehistory
- Weak & Strong AI Hypotheses
- The Very Possibility Of Mechanical Intelligence
- Intentionality
- Symbol Grounding
- The Chinese Room
- Consciousness, Phenomenology, & Qualia
- Social Ramifications Of Artificial Intelligence

Prehistory

- Philosophy of Mind
- Philosophy of Science

Weak Artificial Intelligence

- Weak AI – We will eventually be able to produce machines that act as if they are intelligent.
- Are there things that computers will never be able to do, no matter how we program them?
- Are our techniques for designing intelligent programs inherently flawed?
- Is it the case that, while intelligent programs can exist in principle, the process of constructing them is infeasible?

Strong Artificial Intelligence

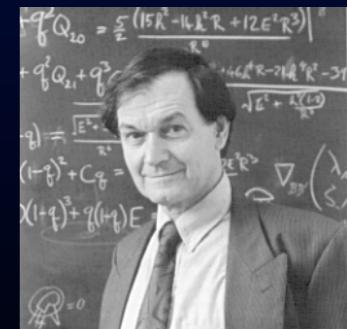
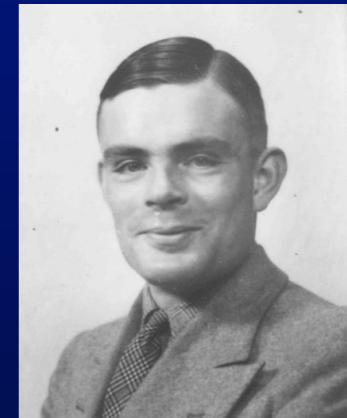
- Strong AI – At least some machines that act intelligently have real conscious minds.

What do you think?

Is Intelligent Behavior Possible?

Will computers ever behave intelligently?

- argument from disability
- argument from informality
 - GOFAI
- the mathematical objection

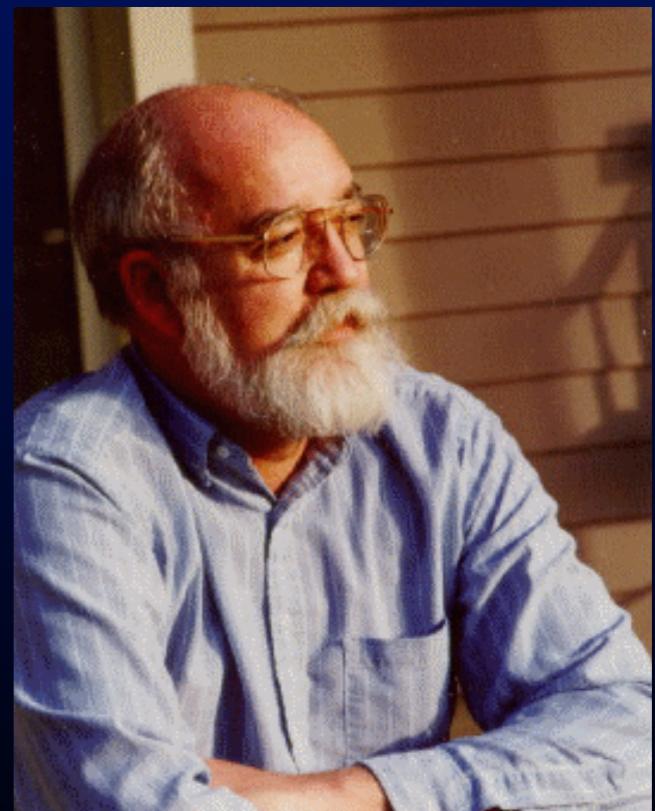


Intentionality

What does it mean for an agent to hold a particular propositional attitude or intentional state (e.g., believing, knowing, desiring, etc.)?

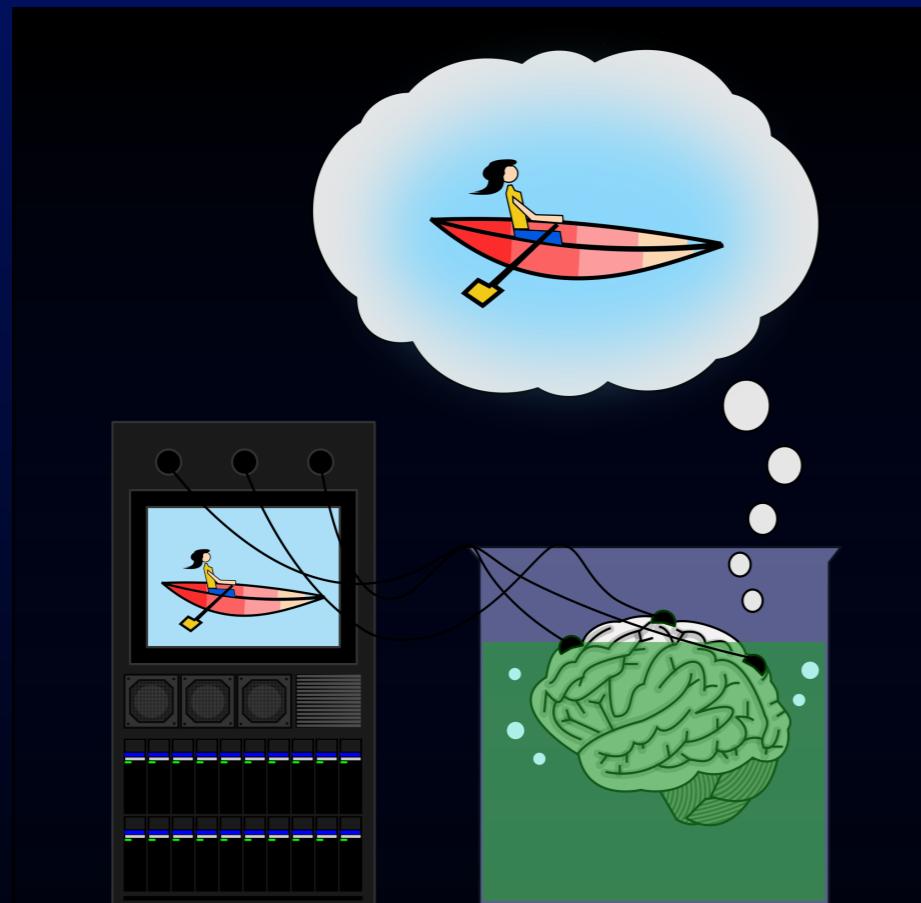
What does it mean to believe or know something?

- intentional stance
- correspondence theory
 - connection to sensory input
 - affects action selection



Symbol Grounding

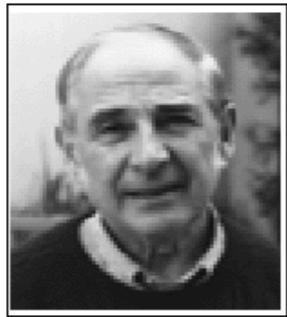
- Internal representations need to have causal semantics.
- wide content versus narrow content



The Chinese Room

3 John Searle, 1980a, 1980b, 1990b

The Chinese Room argument. Imagine that a man who does not speak Chinese sits in a room and is passed Chinese symbols through a slot in the door. To him, the symbols are just so many squiggles and squoggles. But he reads an English-language rule book that tells him how to manipulate the symbols and which ones to send back out. To the Chinese speakers outside, whoever (or whatever) is in the room is carrying on an intelligent conversation. But the man in the Chinese Room does not understand Chinese; he is merely manipulating symbols according to a rule book. He is instantiating a formal program, which passes the Turing test for intelligence, but nevertheless he does not understand Chinese. This shows that instantiation of a formal program is not enough to produce semantic understanding or intentionality. **Note:** For more on Turing tests, see Map 2. For more on formal programs and instantiation, see the "Is the brain a computer?" arguments on Map 1, the "Can functional states generate consciousness?" arguments on Map 6, and sidebar, "Formal Systems: An Overview," on Map 7.



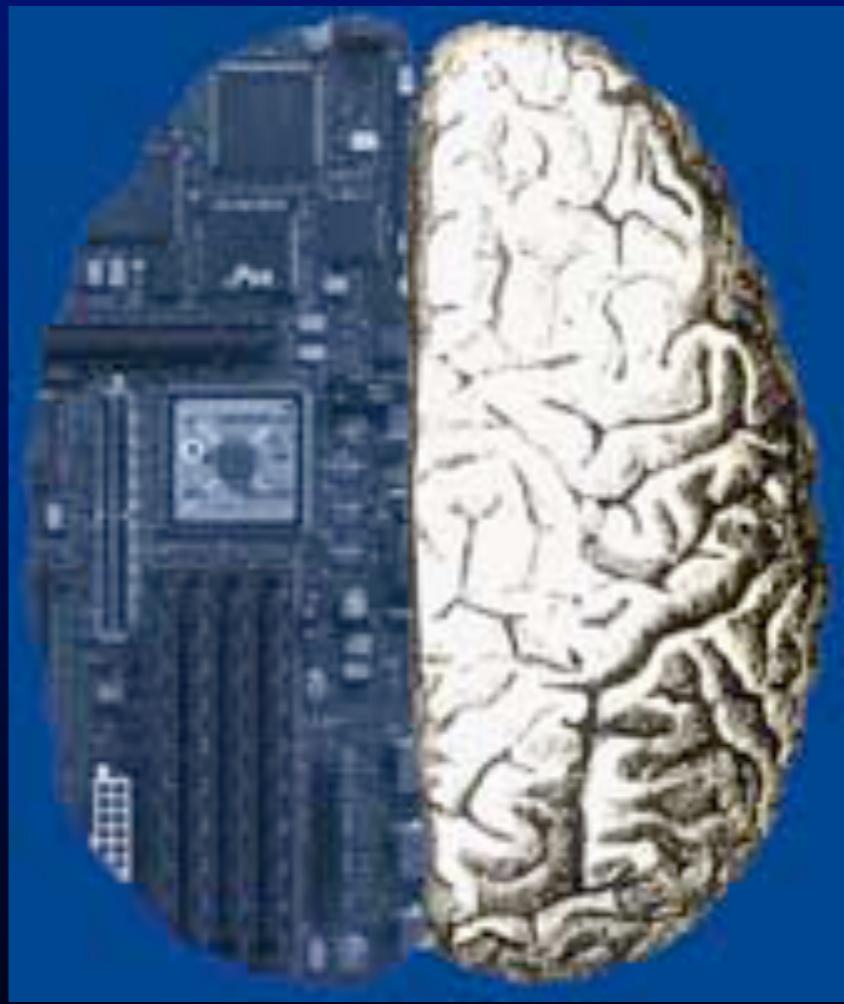
John Searle



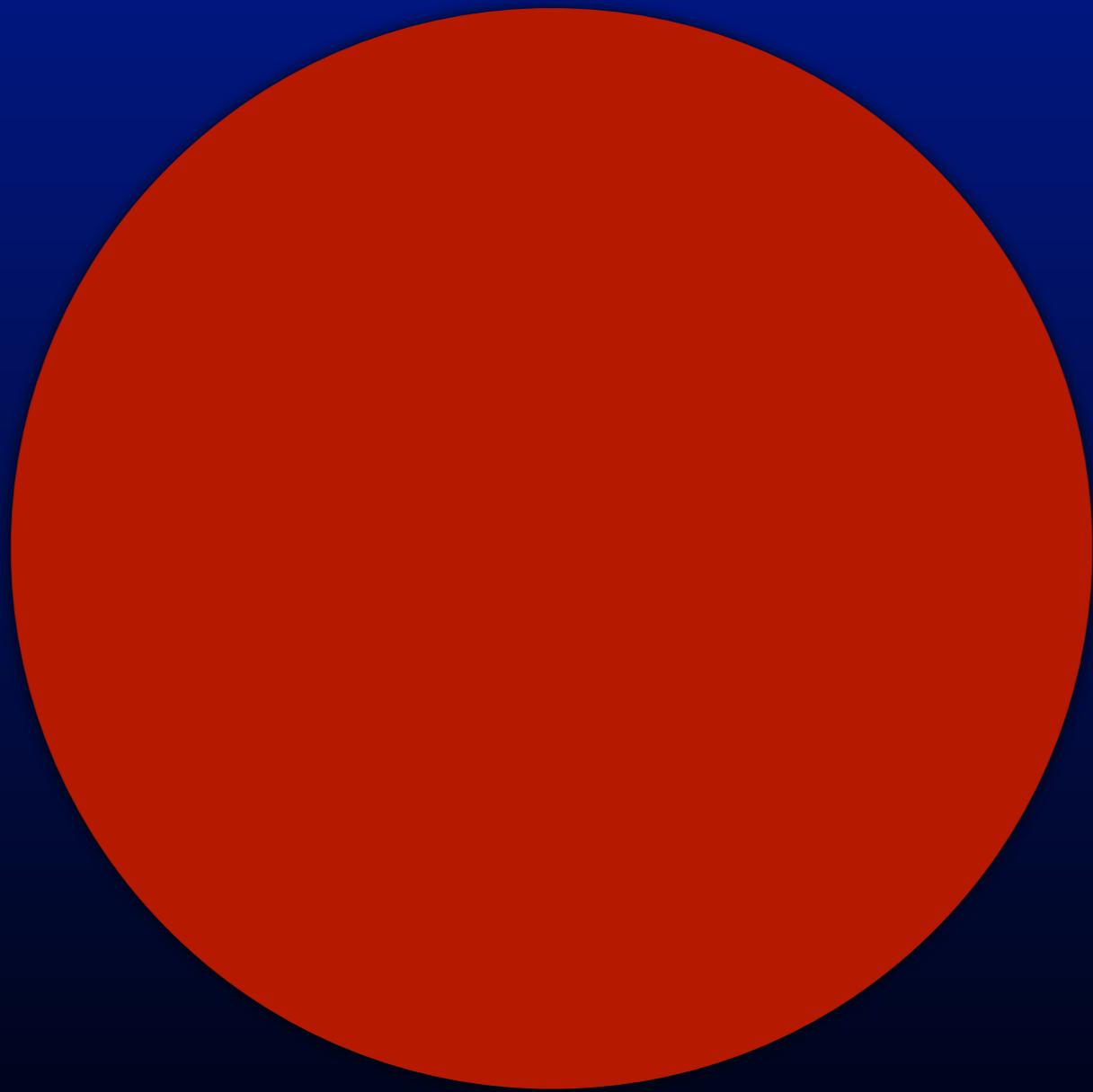
in · ten · tion · al · it · y: The property (in reference to a mental state) of being directed at a state of affairs in the world. For example, the belief that Sally is in front of me is directed at a person, Sally, in the world. Intentionality is sometimes taken to be synonymous with representation, understanding, consciousness, meaning, and semantics. Although there are important and subtle distinctions in the definitions of "intentionality," "understanding," "semantics," and "meaning," in this debate they are sometimes used synonymously.

Consciousness

The Brain Prosthesis Thought Experiment

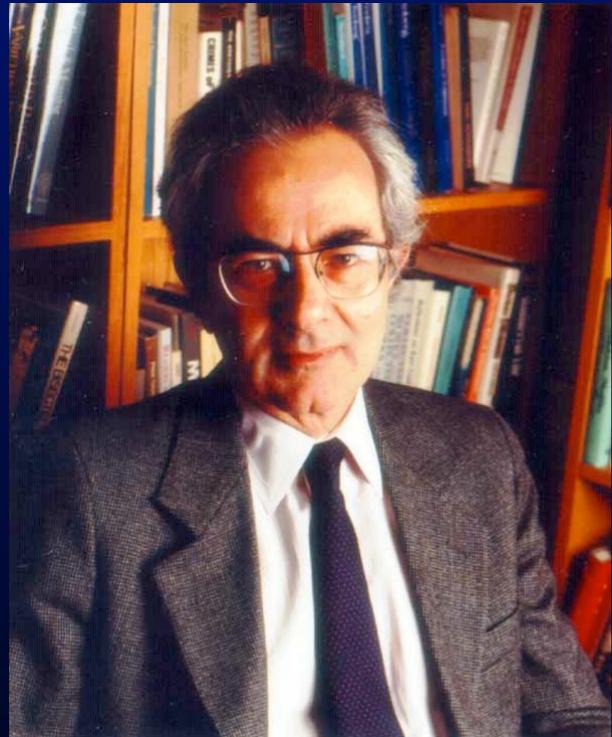


Phenomenology & Qualia



More Qualia

What is it like to be a bat?



Philosophical Zombies



Eliminative Materialism



Ethics & Social Ramifications

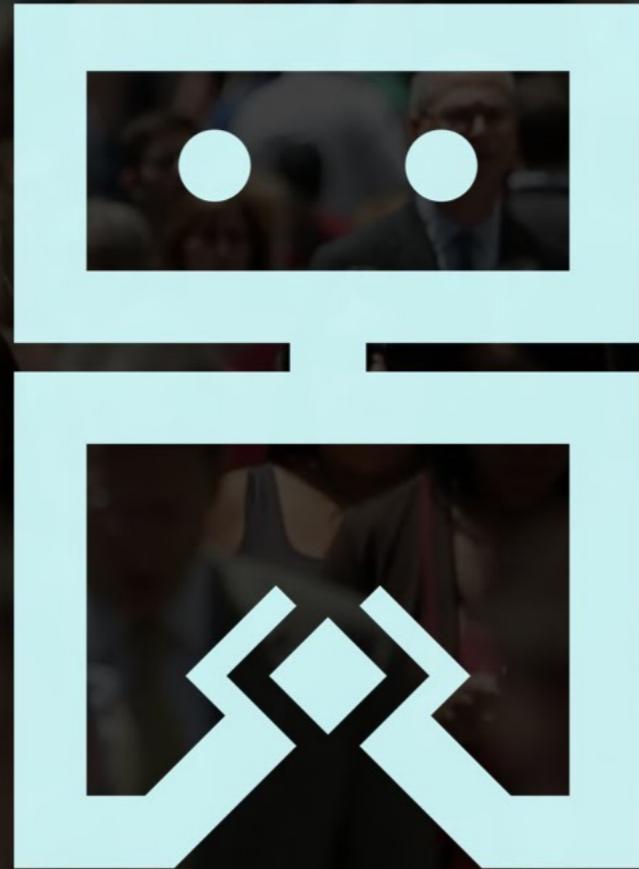
- People might lose their jobs to automation.
- People might have too much (or too little) leisure time.
- People might lose their sense of being unique.
- People might lose some of their privacy rights.
- The use of AI systems might result in a loss of accountability.
- The success of AI might mean the end of the human race.

Ethics & Social Ramifications

- People might lose their jobs to automation.
- People might have too much (or too little) leisure time.
- People might lose their sense of being unique.
- People might lose some of their privacy rights.
- The use of AI systems might result in a loss of accountability.
- The success of AI might mean the end of the human race.

Automation & The Economy

Humans Need Not Apply



Doomsayers & Transhumanists

