Philosophy of Artificial Intelligence

Gilbert Harman

Monday, October 6, 2003

Philosophy

Artifacts

- Dualism
- More Analogies

People as Models for AI

Artifacts as Persons

MORAL?

Philosophy and AI

An important philosophical imperative: “Know thyself!”

What is it to be a person?

What is the difference between people and other animals?

“Man is the rational animal.” Or: man is the animal that thinks.

Inside/Outside

I know what it is like to be me from the inside.

I want to know how what it is like to be me from the inside fits with what it is like to be me from the outside as revealed to other people, or science

Related issue: knowing what it's like to be you via an understanding of you from the outside.

The problem of other minds.

Analogies

Metaphors and analogies provide one way of understanding things.

- Atomism.
- Wave theory of sound.
- Current theory of electricity.
- Particle and wave theories of light.

Argument by analogy to other minds.

Models of Mind

Mechanical toys of the 16th Century suggested a person might be a machine.

Developments in AI suggest models of human intelligence.

Descartes' Dualism

Background physics: contact mechanics, billiard ball models.

- No action at a distance
- No fields: gravity, magnetism, electricity.

Argument that mind is not explainable mechanically.

- Mind involves thought, association of ideas, and reasoning.
- Other animals can perhaps be explained mechanically. But not people. People are basically their minds, attached to their bodies.
- Perhaps minds can survive the destruction of the bodies.
- Explanation of mind is different from explanation of bodies. No mere mechanical explanation of mind.

Conversation

In particular, Descartes argued that it is inconceivable that mechanical principles could explain ordinary human conversation, which has features of novelty and appropriateness.

Compare Turing Test in Alan Turing, "Computing Machinery and Intelligence".

According to the linguist Noam Chomsky, explaining ordinary coversation is a mystery rather than a problem.

Animals

According to Descartes: animals are in principle explicable mechanically.

They do not think or reason in the way people do. They do not act on reasons.

They do not have immortal souls.

They do not have language.

Some followers of Descartes went around kicking dogs in order to show their allegiance to dualism.

Interaction

Mind and body must interact in perception and in action.

This raises the problem: how? How can something that is not a body have an effect on body? And vice versa?

Descartes argues that the point of interaction between the two realms occurs in a certain gland in the brain, the pineal gland.

Developments in Physics

Later developments: changes in physics allowed for nonmechanical effects and action at a distance.

This opens up new possibilities for mind body interaction.

Maybe the effect of mind on body is like the effect of gravity or the effect of magnetism.

Perhaps a mind is something like a field.

Quantum physics suggests additional possibilities.

More Analogies

Flow charts in programming: Psychological theories as flow charts.

Information theory. Mind as an information processing system.

Telephone switchboard analogy.

Computer Theory and Programming

- Logic programming: thinking as theorem proving.
- Post production systems: grammars as production systems, minds as production systems.
- Computer: mind as a computer; person as a computer in a robot.
- Subroutines and modularity, psychological modularity: perceptual systems, language systems, motor systems, face-recognition system.
- Expert systems: intelligence as expertise.
- Pattern recognition and statistical learning theory: psychological learning as pattern recognition.

People as Models for AI

Simulation of ideas from psychology as ideas for CS.

- neural nets → computational systems → connectionism
- paradigm based thinking → nearest neighbor systems
- probability → probability nets

Artifacts as Alive, Conscious, Persons

Artificial intelligence

Artificial life

Artificial consciousness.

Behaviorism

Behaviorism equates mental states and occurrences with behavior.

- Including behavioral tendencies and dispositions.
- Magnetic as an example of a disposition.
- Magnetizable as a second-order disposition.

Turing on computational intelligence

Problem about stoicism.

Double-aspect Theory

Some physical events are mental events from the inside.

Pain is activity in certain C-fibers.

Mind-body identity theory.

- Compare: lightning is an electrical discharge.
- Water is H_2O

“Functionalism”

Mental events can be identified with whatever physical events have the relevant causal properties.

It does not matter what they are made of.

Chinese Room Argument

A system might behave as a speaker of Chinese and contain events with the right functional properties without understanding Chinese.

A computer simulating a Chinese speaker does not understand Chinese.

MORAL?