

# **AITA : Limitations and Misconceptions of AI**

© John A. Bullinaria, 2003

1. Limitations and Misconceptions
2. Lucas and Penrose
3. Dreyfus
4. Searle

## Limitations and Misconceptions

In this lecture you will discuss (rather than be lectured about) two related aspects of AI:

1. **Limitations** – What intelligent things are there that AI can never do?
2. **Misconceptions** – What can AI actually do, that some people think it can not do?

Relevant hypotheses: *weak AI* – that machines can act *as if* they are intelligent, and *strong AI* – that machines which act intelligently are *actually thinking with conscious minds*.

## Lucas and Penrose

*Gödel's Incompleteness Theorem* tells us that any formal system (language plus rules) rich enough to include arithmetic must contain true theorems that cannot be proved within that system. J.R. Lucas, and more recently Roger Penrose, have tried to argue that this implies that human reasoning cannot be mechanised, with the obvious consequent implications for AI. There are several problems with their claims, and very few AI researchers now believe their arguments. (For details and references see: Russell & Norvig, p 949.)

## Dreyfus

Hubert Dreyfus is a philosopher who, over the last thirty years, has written numerous books and papers about things he thinks computers and/or AI cannot do. Learning, uncertainty, common sense knowledge, control of sensors, and so on. He appears to have been proven wrong on all counts. Though, to be fair, many of his criticisms were more aimed at *GOFAI* rather than AI in general. (For details and references see: Russell & Norvig, p 951.)

## Searle

John Searle's claim to fame is his *Chinese Room argument*. Consider a room containing a human who only understands English, a rule book, and various stacks of paper. Chinese words are handed in, and the human uses the rule book and stacks of paper to determine which Chinese words to hand out. He passes the Turing test in Chinese, but has no understanding of Chinese or of how he passed the Turing test. Searle believes this to be of relevance to the possibility of strong AI. Others do not! Compare this with the *Brain Prosthesis Experiment*. (For details and references see: Russell & Norvig, p 958.)

## Overview and Reading

1. One frequently comes across people attempting to criticise AI, or to argue that particular weak or strong forms of AI are not possible. It is generally a waste of time trying to rehearse answers to all these flawed arguments. A good general understanding of AI will be sufficient to deal with such situations.
2. However, it is worth being familiar with some oft quoted cases: Gödel (and his incompleteness theorem), Lucas and Penrose (and their claimed implications of Gödel's theorem), Dreyfus (and his general arguments about what computers supposedly cannot do), and Searle (and his Chinese room argument and the related Brain Prosthesis Experiment).

### Reading

1. Russell & Norvig: Chapter 26.
2. D.R. Hofstadter & D.C. Dennett (1981). *The Mind's I*, Penguin Books.