

PHIL 280

PHILOSOPHY AND COGNITIVE SCIENCE:

Mind from Computation?

Tu, Th 2:00-3:15

Christopher Cherniak

Office: 1103B Skinner

Office Hrs: by appt, &

Tu 3:30-4:30,

Th 3:15-4:00

The computer model of the human mind serves as a central unifying conceptual framework for the cognitive sciences, which lie at the intersection of philosophy, psychology, computer science, and brain science.

This course proceeds from an introduction to computation theory, to some philosophy of mind -- in particular, from the major unsolvability results of computation theory to questions regarding whether machines can (ever) think. The first half of the course is organized around the key concept of computation theory, that of the algorithm or program-schematic. The second half focusses on the philosophical adequacy of computational psychology, the information-processing model of mind; more concretely, what are the possibilities, prospects, and limitations of artificial intelligence?

There are homework problem assignments in connection with the first part of the course (some of them using PC software developed for the course by the UM Philosophy Department). For the second half of the course, the main requirement is three short papers (about 5 pages each: typed, double-spaced, with 1" margins) on assigned, well-defined questions. This is not a programming course, nor does it require any programming background. Students not majoring in philosophy are welcome.

Principal texts (in order of use):

D. Harel, Algorithmics, 3rd ed, 2004.

(A. Dewdney, The (New) Turing Omnibus, 1993.)

R. & D. Cummins, eds, Minds, Brains, and Computers, 2000.

H. Dreyfus, What Computers Still Can't Do, "3rd" ed, 1992.

(W. Lycan, ed, Mind and Cognition, 1990/1999.)

(J. Haugeland, ed., Mind Design II, 2nd ed, 1997.)

MindSet software packages. Downloadable from:

www.glue.umd.edu/~cherniak/philcomp

PCOMP logic engine simulator (TM* & RAM*)

) UM

THERMO parallel computation modeller

) CogSci

Micro-SHRDLU blocks-world natural-language

) Courseware

"understander"

ALICE [2000, 2001, 2004 winner of Loebner "Turing Test" Prize]

www.alicebot.org/ www.20q.net

(ELIZA "psychiatrist" program, J. Weizenbaum)

(RACTER conversationalist program, INRAC Associates)

MIT Encyclopedia of Cognitive Sciences: <http://cognet.mit.edu>

Dictionary of Algorithms: www.nist.gov/dads/

History

<http://plato.stanford.edu/entries/computing-history/>
Stanford Museum: www.computerhistory.org/
Turing: www.turing.org.uk/turing/
Bletchley Park: www.codesandciphers.org.uk/

I. The Nature of Computation

1. The concept of an algorithm: a finite, completely specified set of instructions.
Harel, Part I, especially ch 1
<http://plato.stanford.edu/>
(Minsky, ch 5)
(Kleene, pp. 223-231)
2. Turing machines: a standard format for algorithms.
Harel, ch 9, especially 219-231 [= 223-238, 2nd ed]
TM* PC onscreen Turing machine simulator/interpreter (on PCOMP)
3. The universal Turing machine: an idealized computer that can execute any algorithm.
Harel, ch 9, especially 236-238 [= 242-244]
(Minsky, ch 7)
UTM on TM* on PCOMP
4. Unsolvability: the Halting Problem cannot be solved by any Turing machine.
Harel, ch 8, especially 198-207 [= 203-211]
(Minsky, ch 8)
Kleene, pp. 175-183, 246
(Courant & Robbins, pp. 77-88 (countability & uncountability))
W. Quine, "The Ways of Paradox," in The Ways of Paradox, 1966

[Computational intractability vs absolute unsolvability.]
H. Lewis & C. Papadimitriou, "The Efficiency of Algorithms," Scientific American, Jan 1978
L. Stockmeyer & A. Chandra, "Intrinsically Difficult Problems," Scientific American, May 1979
Harel, ch 7
5. The organization of von Neumann machines: a simple model of a random-access memory computer.
RAM* PC onscreen von Neumann machine simulator/interpreter (on PCOMP)
(Dewdney, chs 17 (& 48)) [Harel ?]
- [6. Parallel computation: out of the von Neumann bottleneck?]
S. Kirkpatrick, C. Gelatt, & M. Vecchi, "Optimization by Simulated Annealing," Science 220, 1983 [Fig 9]
THERMO parallel computation modeller
D. Tank & J. Hopfield, "Collective Computation in Neuronlike Circuits," Scientific American 257, 1987
(Harel, ch 10)

II. The Computer as Model of the Mind

1. Are minds machines: implications of unsolvability theorems?
D. Dennett, "The Abilities of Men and Machines," in Brainstorms
(J. Lucas, "Minds, Machines, and Godel," (1961) www.jstor.org)
Lucas, The Freedom of the Will (1970) chs 24-26
<http://users.ox.ac.uk/~jrlucas/mmg.html>
(R. Penrose, The Emperor's New Mind (1989) ch 4)
2. Can machines think: Turing's test.
Harel, ch 15 [= 12] [[jstor](#)]
A. Turing, "Computing Machinery and Intelligence," in Cummins
ELIZA "psychiatrist" program & RACTER conversationalist
Program
vs M-SHRDLU "blocks-world" natural language understander
(Winograd, ch. 1, Understanding Natural Language; cf Cummins)
3. What are mental states: computational psychology.
H. Putnam, "Robots: Machines or Artificially Created Life?"
1964
(Putnam, "Minds and Machines," in Cummins) - on [jstor](#)
(J. Fodor, "The Mind-Body Problem," Scientific American, Jan
1981)
N. Block, "The Computer Model of the Mind"
C. Cherniak, "The Wager," AI Magazine, Aug 1986 - on homepage
4. Against artificial intelligence: a phenomenological critique.
H. Dreyfus, Introduction to 2nd Edition (1979), What Computers
Still Can't Do; also, Introduction to "3rd" Edition (1992)
(Dreyfus, pp. 155-227, 285-305)
[C. Cherniak, "Undebuggability and Cognitive Science,"
Communications of Association for Computing Machinery
April 1988] www.glue.umd.edu/~cherniak/