



Abacus

Enigma Machine

ENIAC

Apple I (built from kit)

KISMET

Enigma Machine: German cipher machine. The picture shows a 3 rotor model, <http://www.chss.montclair.edu/~pererat/mcpu.htm>

ENIAC: Electrical Numerical Integrator and Computer, Smithsonian Photo #90-7164B, Laurie Minor-Penland, <http://photo2.si.edu/infoage/infoage.html>

Apple I: The original Apple Computer was sold as a kit. Purchasers had to produce their own cases, <http://photo2.si.edu/infoage/infoage.html>

KISMET: Kismet is an autonomous robot designed for social interactions with humans designed by Cynthia Breazeal at MIT's *Sociable Machines Project*.

The Possibility of Artificial Intelligence, SPRING 2007

Phil 355 / Psy 355: T/Th 10:30 – 11:50

Professor William Seeley, 323 Stager Hall

Office Hours: T/Th 12:00 – 1:00 & by appointment

william.seeley@fandm.edu

Course Description:

What are minds? Are minds like computers? If so, how much is a computer like a mind? Is it possible for a machine or a computer to think? If it were possible, what would these thoughts be like? Would they be just like ours? Could machines feel emotions or make genuinely moral choices? Do these questions have any bearing on the possibility of artificial intelligence? Underlying all these issues is a central question concerning the nature of thinking itself. Before we can know what sorts of things can be thinkers we must first come to an agreement on what it is to be a thinker at all.

It was once thought that thinking, the ability to form rational thoughts, was what set human beings from the rest of the universe. Today cognitive scientists look to commonalities between minds and computers to help them understand complex cognitive tasks like visual perception, rationality, and language development. In this course we will examine and evaluate some contemporary concepts and issues in philosophy, psychology, and computer science that are critical to both research in artificial intelligence and the computational theory of mind. The topics covered in this course will include: intentionality, representation, consciousness, the computational theory of mind, artificial life, and alternatives to the traditional (GOFAI) model for AI.

Course Goals:

The goals of this course are threefold. We will try to come to an understanding of what artificial intelligence is, and more importantly what it is not, as a research program in psychology and computer science. In this context we will evaluate the validity of several philosophical problems associated with artificial intelligence. This debate will be used to evaluate the traditional GOFAI model and introduce contemporary alternatives.

Texts:

- *Mindware*, Andy Clark (New York: Oxford, 2001). **MW**
- *Mind Design II*, ed. John Haugeland (Cambridge, MA: MIT Press, 1997). **MD**
- *What Computers Still Can't Do*, Hubert Dreyfus (Cambridge, MA: MIT Press, 1993). **WCSCD**
- *Flesh and Machines*, Rodney Brooks (New York: Vintage Books, 2002). **FM**
- Electronic resources and pdf files on eDisk. **pdf**

Requirements:

- Two 3 page response papers due February 19th & April 15th.
- A 6 page paper on an assigned topic due at the midterm. Students will be given a choice between two topics that require them to synthesize the material from the first half of the semester.
- XOR Exercise: a set of exercises designed by Kim Plunkett & Jeffery Elman to teach students about connectionist systems.
- An 8 - 10 page final paper on a topic of your choosing (50%). All students must clear final paper topics with me before April 1st. The final paper is due at the first day of exams.
- Late papers will NOT be accepted!

Schedule of Readings:

Topic 1 : Introduction: What is AI?

01/24	Brooks	"We Are Special/We Are Not Special" <i>FM</i>
01/29	Clark: Haugeland: Copeland: Intelligent By Design: Kismet: Play with ELIZA AI Topics Homepage	Chapter 1, "Meat Machines: Mindware as Software" <i>MW</i> "What Is Mind Design?" 1-28 <i>MD</i> Chapter 2, "Some Dazzling Exhibits" <i>pdf</i> http://www.pbs.org/saf/1303/features/AI.htm http://www.ai.mit.edu/projects/sociable/ongoing-research.html http://www.manifestation.com/neurotoys/eliza.php3 http://www.aaai.org/aitopics/

Topic 2: Minds, Patterns, and Representations

01/31	Clark: Dennett:	Chapter 3, "Patterns, Contents, & Causes" <i>MW</i> "True Believers: The Intentional Strategy and Why It Works" <i>MD</i>
02/05 - 02/07	Clark: Turing: Newell & Simon	Chapter 2, "Symbol Systems" <i>MD</i> "Computing Machinery & Intelligence" <i>MD</i> "Computer Science as Empirical Enquiry: Symbols & Search" <i>MD</i>
02/12	Searle: Copeland:	"Minds, Brains, & Programs" <i>MD</i> "The Chinese Room from a Logical Point of View" <i>pdf</i>
02/14	Discussion Day	Is the symbol system hypothesis a good criteria for intelligence?
02/19	Short paper (1000) words:	(See Topics Below)

Topic 3: The Frame Problem:

02/19 02/21	Minsky: Dreyfus:	"A Framework for Representing Knowledge" <i>MD</i> "Introduction" (From Micro-Worlds to Knowledge Representation) <i>WCSCD</i>
----------------	---------------------	---

Topic 4: AI at an Impasse?

02/26	Dreyfus:	"Assumptions Underlying Persistent Optimism" <i>WCSCD</i>
02/28	Discussion day:	Is the classical GOFAI model for AI doomed to failure? Mid-term Paper Assigned (topic TBA)

Topic 5: Connectionism

03/04	Clark: Rumelhart: Plunkett & Elman:	Chapter 4, "Connectionism" <i>MW</i> "The Architecture of Mind: A Connectionist Approach" <i>MD</i> "The Methodology of Simulations" <i>pdf</i>
03/06	Churchland: Rosenberg:	"On the Nature of Theories: A Neurocomputational Perspective" <i>MD</i> "Connectionism & Cognition" <i>MD</i>
03/11 03/13	Fodor & Pylyshyn: Clark	"Connectionism and Cognitive Architecture: A Critical Analysis" <i>MD</i> "The Presence of a Symbol" <i>pdf</i>
02/28 - 03/13	Connectionist Modeling: Plunkett & Elman	"Learning to use the simulator" (Chapter 3) <i>pdf</i> "Learning internal representations" (Chapter 4) <i>pdf</i>

Topic 6: Challenges to the Traditional Model

Heideggerian AI:

03/25 Dreyfus: "Alternatives to the Traditional Assumptions" **WCSCD**

03/27 Preston "Heidegger and Artificial Intelligence" **pdf**

Robotics & A-Life:

04/01 Clark: Chapter 6: "Robotics & Artificial Life" **MW**
Brooks: "Intelligence without Representation" **MD**

04/03 Kirsch "Today the Earwig, Tomorrow Man?" **pdf**

Dynamics:

04/08 Clark: Chapter 7: "Dynamics" **MW**
van Gelder: "Dynamics & Cognition" **MD**

04/10 Markman & Dietrich "Extending the Classical View of Representation" **pdf**

04/15 Discussion Day Can Heideggerian Approaches to AI model higher level cognition?

Affective Computing:

04/17- Picard "Envisioning Affective Computing" **AC**
04/22

04/24 Prinz "Emotions" **pdf**
AI& Emotions <http://www.inf.ed.ac.uk/events/hotseat/aboutAI.html>

04/29 Discussion Day Do models for affective computing help us model the role of emotions in reasoning?

Some Miscellaneous Notes and Guidelines:

Moral behavior is the grounds for, and the framework of, a healthy society. In this regard it is each of our responsibility as an individual within the community of our classroom to act responsibly. This includes following the rules and guidelines set out by the College for academic behavior. Plagiarism is a serious matter. It goes without saying that each of you is expected to do his or her own work and to cite EVERY text that is used to prepare a paper for this class. As a general rule, I ask that you not use the internet for your research except as assigned in class.

The reading list for this class is arranged in topics as opposed to individual sessions. I will announce the particular readings for each class as we go along. This will allow us some flexibility in discussion so that we can spend more time on issues of interest to the class. I reserve the right to make changes to the syllabus as we go along in order to accommodate class interests. I will also occasionally upload supplementary materials to eDisk for students interested in pursuing particular issues beyond class discussion.

ASSIGNMENTS:

Paper Topic #1: The purpose of these *response papers* is for you to evaluate a standard argument in the literature. This is not an *opinion paper* as standardly construed. The method of philosophy is critical analysis. We are interested in understanding the reasons behind values and beliefs, or better, the reasons that provide rational support for the beliefs that we hold. These reasons, if good, ought to provide logical support for our values and beliefs. In the following paper you should: identify the theoretical problem at hand; rehearse the standard argument for the position identified; & evaluate whether these reasons genuinely support that position. The first step identifies the problem space that you are addressing. The second step should have the form of a *rational reconstruction*. In a rational reconstruction one does their best to give an argument a fair shake. You should do your best to make the standard argument as plausible as you can. The final step is to respond. Your response should identify a step in the standard argument that you find to be in error. In this case it could be a step in the Chinese Room Argument itself or in Copeland's critique of its logical structure. The key here is that you are not arguing for the truth or falsity of the target position per se. Rather you are arguing that the reasoning offered does not suffice to establish that position as a conclusion.

- a: **Is there good reason to believe that the Chinese room really can't understand Chinese?**
- b: **Discuss the logic of the Chinese Room Argument. Is it sound?**

Due Date: February 19

Connectionist Modeling Assignment: Complete the exercises in chapters 3 & 4 of Plunkett & Elman (*pdf*). This will require you to download the Tlearn software from Elman's website. I will introduce you to the software and go over the assignment in class on March 4th. The purpose of this set of assignments is to familiarize you with the structure of connectionist architectures and the nature of connectionist models of cognitive processes. I will provide you with the answer sheets so that you can check your work (these can also be used to help work out solutions if you get stuck-top-down problem solving strategies are often useful for learning). You will be required to hand in print-outs of the exercises to verify that you have done them.

Due Date: March 14

Midterm Assignment: Write a 6 page paper on one of two topics to be distributed at the midterm (March 13). Your paper should be double-spaced in 12 point font with 1" margins. The purpose of this paper is twofold: a) identify & evaluate a standard argument in the literature; and b) demonstrate that you can synthesize the diverse range of material covered in the first half of the semester into a coherent position.

- a) **Are connectionist models of AI genuinely alternatives to GOFAI models? Make sure in your answer to discuss Marr's three levels of analysis, differences between connectionist and GOFAI architectures, and the role played by linguistic behavior in evaluations of both types of models.**
- b) **Use the discussion of either *the psychological hypothesis* (as exemplified by Fodor) or *the metaphysical hypothesis* (as exemplified by Minsky) but not both to evaluate Dreyfus' criticism of the GOFAI model for artificial intelligence.**

Topics Distributed: 03/13
Due Date: 04/01

Paper Topic #3:

- a: **Are the alternatives to GOFAI (e.g., Robotics/Alife, Dynamic Systems, & Heideggerian AI) genuinely anti-representational?**

Due Date: April 15

Final paper: there is no final exam; your final paper (8-10 double-spaced pages) is due on the scheduled exam date; you must clear your topic with me by April 1st

Bibliography:

Topic 1: Introduction: What is AI?

- Rodney Brooks, "We Are Special/We Are Not Special," *Flesh and Machines: How Robots Will Change Us* (New York: Vantage), 2002, 148-196.
- Andy Clark, "Chapter 1: Meat Machines: Mindware as Software," *Mindware* (New York: Oxford, 2001), 7-27.
- Jack Copeland, "Chapter 2: Some Dazzling Exhibits," *Artificial Intelligence* (Malden, MA: Blackwell Publishers, 1993), 11-32.
- John Haugeland, "What Is Mind Design?" in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 1-28.

Supplemental:

- Daniel Dennett, "Artificial Intelligence as Philosophy and Psychology," *Brainstorms* (Cambridge, MA: MIT Press, 1993), 109-128.

Topic 2: Minds, Patterns, and Representations

- Andy Clark, "Chapter 3: Patterns, Contents, & Causes," *Mindware* (New York: Oxford, 2001), 43-61.
- Daniel Dennett, "True Believers: The Intentional Strategy and Why It Works," in ed. John Haugeland (Cambridge, MA: MIT Press, 1997), 57-80.
- Andy Clark, "Chapter 2: Symbol Systems," *Mindware* (New York: Oxford, 2001), 28-42.
- Alan Turing, "Computing Machinery & Intelligence," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 29-56.
- Allen Newell & Herbert A. Simon, "Computer Science as Empirical Enquiry: Symbols & Search," in ed. John Haugeland, *Mind Design II*, (Cambridge, MA: MIT Press, 1997), 81-110.
- John Searle, "Minds, Brains, & Programs," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 183-204.
- Jack Copeland, "The Chinese Room from a Logical Point of View," in eds. John Preston & Mark Bishop, *Views into the Chinese Room* (New York: Oxford University Press, 2002), 109-122.

Supplemental:

- Jack Copeland, "Chapter 4: The Symbol System Hypothesis," *Artificial Intelligence* (Malden, MA: Blackwell Publishers, 1993), 58-82.

Topic 3: The Frame Problem:

- Marvin Minsky, "A Framework for Representing Knowledge," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 111-142.
- Hubert Dreyfus, "Introduction" *What Computers Still Can't Do* (Cambridge, MA: MIT Press, 1993), 1-66.

Supplemental:

- Daniel Dennett, "Cognitive Wheels: The Frame Problem of AI," in ed. Hookway, C, *Minds, Machines, and Evolution: Philosophical Studies* (Cambridge, UK: Cambridge University Press, 1984) 147-170.

Topic 4: AI at an Impasse?

- Hubert Dreyfus, "Assumptions Underlying Persistent Optimism," *What Computers Still Can't Do* (Cambridge, MA: MIT Press, 1993), 155-227.

Topic 6: Connectionism?

- Andy Clark, "Chapter 4: Connectionism," *Mindware* (Cambridge, MA: MIT Press, 2001), 62-83.
- David Rumelhart, "The Architecture of Mind: A Connectionist Approach," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 205-232.
- Paul Churchland, "On the Nature of Theories: A Neurocomputational Perspective," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 251-292.
- Jay Rosenberg, "Connectionism & Cognition," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 293-308.
- Jerry Fodor & Zenon Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 309-350.
- Andy Clark, "The Presence of a Symbol," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 377-393.

Simulation Project:

- Plunkett & Elman, "The Methodology of Simulations," *Exercises in Rethinking Innateness* (Cambridge, MA: MIT Press, 1997), 1-30.
- Plunkett & Elman, "Learning to use the simulator" (Chapter 3), *Exercises in Rethinking Innateness* (Cambridge, MA: MIT Press, 1997), 31-74.
- Plunkett & Elman, "Learning internal representations" (Chapter 4), *Exercises in Rethinking Innateness* (Cambridge, MA: MIT Press, 1997), 75-98.

Supplemental:

- Paul Smolensky, "Connectionism and the Language of Thought," in eds. Macdonald and Macdonald, *Connectionism* (Malden, MA: Blackwell, Publishers, 1995), 164-198

Topic 7: Challenges to the Traditional Model

Heideggerian AI:

- Hubert Dreyfus, "Alternatives to the Traditional Assumptions," *What Computers Still Can't Do* (Cambridge, MA: MIT Press, 1993), 231-282.
- Beth Preston, "Heidegger and Artificial Intelligence," *Philosophy and Phenomenological Research*, 53(1), 1993, 43-69.

Robotics and A-Life

- Andy Clark, "Chapter 6: Robotics & Artificial Life," *Mindware* (Cambridge, MA: MIT Press, 2001), 103-119.
- Rodney Brooks, "Intelligence without Representation," in ed. John Haugeland, *Mind Design II* (Cambridge, MA: MIT Press, 1997), 395-420.
- David Kirsch, "Today the Earwig, Tomorrow Man?" *Artificial Intelligence*, 47, 1991, 161-184.

Supplemental:

- Brian Smith, "The Owl and the Electric Encyclopedia," *Artificial Intelligence*, 47, 1991, 251-288.

Dynamics:

- Andy Clark, "Chapter 7: "Dynamics," *Mindware* (Cambridge, MA: MIT Press, 2001), 120-139.
- Timothy van Gelder, "Dynamics & Cognition," *Mind Design II* (Cambridge, MA: MIT Press, 1997), 421-450.
- Arthur Markman & Eric Dietrich, "Extending the Classical View of Representation," *Trends in Cognitive Science*, 4(12), 470-475.

Supplemental:

- Arthur Markman & Eric Dietrich, "In Defense of Representation," *Cognitive Psychology*, 40, 2000, 138-171.

Affective Computing:

- Rosalind Picard, "Envisioning Affective Computing," *Affective Computing* (Cambridge, MA: MIT Press, 1997), 19-137.
- Jesse Prinz, "Emotions," in ed. Paul Thagard, *Handbook of the Philosophy of Psychology & Cognitive Science* (Dordrecht: Elsevier, 2006).

Supplemental:

- Rosalind Picard, "Building Affective Computing," *Affective Computing* (Cambridge, MA: MIT Press, 1997), 139-246.
- Gene Ball and Jack Breese, "Emotion and Personality in a Conversational Agent," in eds. Justine Cassell, Joseph Sullivan, Scott Prevost, & Elizabeth Churchill, *Embodied Conversational Agents* (Cambridge, MA: MIT Press, 2000), 198-219.
- Dylan Evans, 'Can Robots Have Emotions?', *Psychology Review*, 11(1) (September 2004), 2-5.

The Possibility of Artificial Intelligence: syllabus

Date	Readings	Assignments
01/22	Introduction	
01/24	Brooks, "We are/not special"	
01/29	Clark, "Meat machines" Haugeland, "What is mind design" Copeland, "Some dazzling exhibits"	Why aren't ELIZA/KISMET intelligent? (prepare a 1 page answer)
01/31	Clark, "Patterns, causes, & contents" Dennett, "True believers"	
02/05-07	Clark, "Symbol Systems" Turing, "Computing machinery & intelligence" Newell & Simon, "Computer science as empirical enquiry"	
02/12	Searle, "Minds, brains, programs" Copeland, "The Chinese Room from a logical point of view"	
02/14	Discussion Day: Is the symbol system hypothesis a good criteria for intelligence?	
02/19	Minsky, "A framework for representing knowledge"	Paper #1 Due
02/21	Dreyfus, "WCSCD, Introduction to the 1979 Edition"	
02/26	Dreyfus, "Assumptions underlying persistent optimism"	
02/28	Discussion Day: Is the GOFAI model for AI doomed to failure?	
03/04	Clark, "Connectionism" Rumelhart, "The architecture of mind: a connectionist approach" Plunkett & Elman, "The methodology of simulations"	
03/06	Churchland, "On the nature of theories" Rosenberg, "Connectionism & cognition"	
03/11	Fodor & Pylyshyn, "Connectionism & cognitive architecture"	
03/13	Clark, "The presence of a symbol"	
03/04-13	Plunkett & Elman, Chapters 1, 3-4	Exercises from chapters 3 & 4 due Mid-term topic assigned
03/14-24	Spring Break	
03/25	Dreyfus, "Alternatives to the traditional assumptions"	
03/27	Preston, "Heidegger & artificial intelligence"	
04/01	Clark, "Robotics & A-Life" Brooks, "Intelligence without representation"	Mid-term paper due
04/03	Kirsch, "Today the earwig, tomorrow man?"	
04/08	Clark, "Dynamics" van Gelder, "Dynamics & cognition"	
04/10	Markman & Dietrich, "Extending classical views of representation"	
04/15	Discussion Day: Can Heideggerian Approaches to AI model higher level cognition?	Paper #3 Due
04/17	Picard, "Envisioning affective computing"	
04/22	Picard, "Envisioning affective computing"	
04/24	Prinz, "Emotions" Evans, "Can Robots Have Emotions?"	
04/29	Discussion Day: Do models for affective computing help us model the role of emotions in reasoning?	
05/01	Discussion Day: Classical models, alternative models, higher level cognition, & the future of AI	
Exam Week		Final paper due "day of exam"