Philosophy of Artificial Intelligence: A Course Outline

WILLIAM J. RAPAPORT

State University of New York, Buffalo

In the Fall of 1983, I offered a junior/senior-level course in Philosophy of Artificial Intelligence, in the Department of Philosophy at SUNY Fredonia, after returning there from a year's leave to study and do research in computer science and artificial intelligence (AI) at SUNY Buffalo. Of the 30 students enrolled, most were computer-science majors, about a third had no computer background, and only a handful had studied any philosophy. (I might note that enrollments have subsequently increased in the Philosophy Department's AI-related courses, such as logic, philosophy of mind, and epistemology, and that several computer-science students have added philosophy as a second major.) This article describes that course, provides material for use in such a course, and offers a bibliography of relevant articles in the AI, cognitive science, and philosophical literature.

Course Organization and Texts

My original intention, inspired in part by Moulton and Voytek 1980, was to spend the first half of the semester discussing working AI programs, in order to get the students excited about what computers could do, using the first edition of Patrick Henry Winston's *Artificial Intelligence* (Winston 1977) as the text. During the second half of the semester, my plan was to discuss Hubert L. Dreyfus's *What Computers Can't Do* (Dreyfus 1979), as a sort of antidote and source of philosophical reflection.

While I still think that this might have been a good course, I chose instead to use John Haugeland's anthology, *Mind Design* (Haugeland 1981), for the philosophical half of the course, since it offers a wider—and fairer—spectrum of philosophical essays, besides including the most important section of Dreyfus's book. (An annotated table of contents of Haugeland 1981 is given in Appendix A.) The first edition of Winston's AI text (unfortunately now out of print) is far superior to the second edition (Winston 1984) for a course such as this, since it approaches AI by discussing how particular AI programs work, and it includes a LISP primer. There are several other up-to-date AI texts on the market—including Rich 1983, and Charniak and McDermott 1985[1]—but none, to my knowledge, take the approach of Winston 1977.

I might also mention here that excellent sources of background material on AI are Barr *et al.*, eds., *The Handbook of Artificial Intelligence* (1981-1982); Webber and Nilsson, eds., *Readings in Artificial Intelligence* (1981); and Shapiro, ed., *The Encyclopedia of Artificial Intelligence* (forthcoming). The Webber and Nilsson collection has an "Advanced Topics" section with philosophically relevant articles, and the Shapiro volume has several good articles on the relationship between philosophy and AI.[2]

As things turned out, we spent rather more time than planned looking at AI programs (as well as some in-class instruction in LISP) and, so, had less time than planned to read Haugeland's book. Nevertheless, the class discussions were high in philosophical content, much of it initiated by the students. Consequently, it is clear to me now that a better format would be to have the students read and discuss relevant essays from Haugeland while studying the AI programs. Thus, the syllabus presented below represents an idealized, but workable, version of the course.

In addition to short writing assignments throughout the semester (some of which are described below), there was a take-home, essay-question midterm and a term project. For the project, the students had the option of writing a 10-15-page term paper or doing an AI or LISP programming project. These are described in Appendix B.

LISP and AI Programs

LISP is the "language of choice" for AI programming. Fully half of the students chose to learn LISP as their term project, including several who had no prior programming experience. (LISP can, and arguably ought to, be one's first programming language.) Good introductions to what LISP is can be found in Friedman 1974, Hofstadter 1983abc, and Wand 1984. We were fortunate to be able to use the manuscript of an excellent self-tutoring text (Shapiro, 1986). An inexpensive alternative is, as mentioned, the second half of Winston 1977 (which is *not* in the second edition). Other good texts are Winston and Horn 1984 (although I find their programming style very non-LISP-like), Touretzky 1984, and Wilensky 1984.

The other popular AI programming language is PROLOG, whose syntax is like that of first-order logic. But most famous AI programs have been written, and are available, in LISP.

Indeed, if the computing facilities at your institution include a good version of LISP (e.g., Franz LISP, Common LISP, MACLISP, INTERLISP, Portable Standard LISP), then you can—with, perhaps, the assistance of some computer buffs or hackers among your students—implement several AI programs, thus giving the students some "hands on" experience. Manageable versions of SAM (a program that "understands" stories about stereotypical activities, using "scripts"), PAM (a program that "understands" stories involving plans and goals), POLITICS (a program that simulates a political belief system), TALE-SPIN (a program that "creates" stories), and ELI (a natural-language parser) are given in Schank and

Riesbeck 1981. Versions of SIR (a deductive question-answering system) and a number of other AI programs and techniques are given in Shapiro 1979 (out of print, but available in many libraries). And an excellent version (in BASIC) of ELIZA (a simulated Rogerian psychotherapist that converses using unrestricted natural-language input; see below) is available from Artificial Intelligence Research Group, 921 N. La Jolla Ave., Los Angeles, CA 90046.

An Artificial IQ Test

I began the course by giving the students an "Artificial IQ" test: a series of questions not unlike those appearing on "intelligence" tests—all of which can be handled (albeit separately) by various AI programs (though the students were not told this till after they took the test)—together with some discussion questions. The students had several days to work on this before it was discussed in class (Topic 4, below). The test is given in Appendix C and described in this section. I have found this "test" to be an excellent discussion-starter, and I have also used it successfully for "stand-alone" talks on AI.

The first two questions concern Thomas Evans's ANALOGY program (Evans 1968), which reasons about simple geometric analogies using the same strategies that people do. One interesting question that can be raised in this context is this: If (and this is a big assumption) the ability to do such analogies is a mark of intelligence (after all, such questions do appear on real IQ tests, for whatever that might be worth), then is Evans's program (or a computer running the program, or a person following the program) intelligent? On the other hand, if a mere program can solve these problems, in a strictly mechanical fashion, then perhaps this ability is *not* a mark of intelligence.

Question 3 presents computer/human interactions from BORIS (Program "B") and ELIZA (Program "D"). BORIS (Dyer 1982, Lehnert 1983) is a program descended from SAM and PAM that can read fairly complex stories (on topics that it "knows" something about) and answer fairly difficult questions about them. ELIZA is Joseph Weizenbaum's celebrated natural-language program (Weizenbaum 1966, 1976). Unlike BORIS, it can (appear to) converse on any topic whatsoever, although it only has a pattern-matching mechanism and so can hardly be said to understand natural language. (Details on its workings can also be found in Winograd 1983; see Topic 13 A, below.) While most students tend to feel that BORIS is more "intelligent," a few brave souls will point out that ELIZA is more versatile (and fairly accurately simulates a Rogerian psychotherapist) and that, hence, perhaps, it is the more intelligent one. (Cynics will say that it merely shows that Rogerian psychotherapists are not intelligent!)

Question 4 is a problem that can be solved by Earl Sacerdoti's NOAH planning program (Sacerdoti 1977). It calls for fairly sophisticated problem-solving (i.e., programming) skills, involving the use of top-down design and stepwise refinement (i.e., recursively analyzing problems into subproblems that are easier to solve).

Question 5 is adapted from Patrick H. Winston's program that "learns" what an 'arch' is by being shown positive (A, D) and negative (B, C) instances (cf. Winston 1977: 29-43). I use 'CRDL' (pronounced "criddell") instead of 'arch' so as not to beg any questions about what, if anything, is being "learned." That is, if such a program *does* learn something (an open question whose answer depends on the meaning of 'learn,' of course), does it learn what an arch "really" is, or does it "merely" learn how to associate certain symbols with certain diagrams? (Of course, that may be precisely what it means to learn what an arch is.)

Questions 6-10 concern issues in natural-language processing: The sentence in question 6 is (at least) three-ways ambiguous (see Hendrix and Sacerdoti 1981). The point is that a natural-language-understanding program must be able to recognize all three ways and decide which one is appropriate in a given context. The story in question 7 illustrates problems of pronominal reference, another important issue in natural-language understanding. The story in question 8 illustrates the fact that many aspects of stories can be predicted because they are stereotypical: Most students will predict that the article goes on to discuss the role of women in the fire department, to provide statistics, to say that women were not expected to pass, to discuss issues in the feminist movement, etc. The stories in questions 9 and 10 can be handled by SAM and PAM, respectively.

Course Outline

What follows is an outline of the course, together with suggested readings and some writing assignments. The "topic" numbers do not refer to days or weeks; they are merely sequential. The starred readings are central; the others, optional. 'H' refers to Haugeland 1981; 'W' refers to Winston 1977.

Topic 1. Brief discussion of the nature of AI and of cognitive science (an inter- or multidisciplinary study of the nature of the mind and cognition; cf. Gardner 1985), and of philosophy's role. Good background reading for the instructor include Dennett's "Artificial Intelligence as Philosophy and as Psychology" (1979); Churchland's *Matter and Consciousness* (1984), Ch. 6; Ringle's *Philosophical Perspectives in Artificial Intelligence* (1979), and Haugeland 1985 (see note 2). Hand out the AIQ Test.

Topic 2. Issues in the Philosophy of Mind: A brief description of classic theories about the mind-body problem (especially functionalism) and the problem of other minds. Roderick Chisholm's cartoon from Taylor 1983: 16 can serve as a nice outline.

Readings: *Fodor, "The Mind-Body Problem," (1981);
*Dennett, "Intentional Systems," in H;
Putnam, "Reductionism and the Nature of Psychology," in H;
Churchland, *Matter and Consciousness* (1984).

Topic 3. The Turing Test. Presentation and discussion of Turing's thesis that (roughly) the question, "Can a machine think?", should be replaced by the question, "Can a machine fool a human into thinking that s/he is conversing with another human, rather than with the machine?".

Readings: *Turing, "Computing Machinery and Intelligence," (1950);
Hofstadter, "The Turing Test," (1981);
Dennett, "Reflections," (1981);
Davidson, "The Material Mind," in H.

Assignment: Write a 1-2 page *description* of the Turing test, *plus either* (a) a *description* of one of the essays about the Turing test in Hofstadter and Dennett's *The Mind's I* (1981) *or* (b) your reply to the following objection: "According to Turing, a computer can think if you *think* it can think; but you may be mistaken," *or* (c) *your* objections to the Turing test, together with your guess at how Turing would have responded.

(For a discussion of the rationale behind this assignment, see Rapaport 1984ab.)

Topic 4. Discussion of the AIQ test.

Topic 5. Definitions of AI. There are several interestingly different ones; a sampling is presented in Appendix D. Most of these have the form: "AI is the branch of computer science concerned with algorithms that are, mimic, simulate, emulate, model (human) intelligence." Thus, one way to phrase AI's main issue is: Is intelligence algorithmizable (i.e., computable)? Another is: Are mental states and processes (representable by) computable (i.e., recursive) functions?

Readings: *Haugeland, "Semantic Engines," in H;
*Weizenbaum, *Computer Power and Human Reason* (1976), Chaps. 2-3;
*Hofstadter, "Mind, Body and Machine," (1983d);
Handout on Turing machines, e.g., from Schagrin *et al.*, *Logic: A Computer Approach*, Appendix B, or Putnam, "Minds and Machines" (1960);
In response to the students' requests for more information on algorithms, I also handed out Davis's *Computability and Unsolvability* (1958): xv-xviii.

Topic 7. Brief history of AI.

Readings: *W, Chap. 1;
McCorduck, *Machines Who Think* (1979).

Topic 8. The ANALOGY program.

Readings: *W, Chap. 2 (to p. 28):
Evans, "A Program for the Solution of Geometric-Analogy Intelligence Test Questions" (1968): longer and more detailed than Winston's presentation, but, in many ways, much clearer;
Dreyfus, *What Computers Can't Do* (1979): 137ff.

Topic 9. Winston's arch-learning program.

Readings:*W, Chap. 2 (pp. 29-44);
*Cohen and Feigenbaum, *Handbook of AI*, Vol. 3 (1982): 392-96;
Dreyfus, *What Computers Can't Do* (1979): 22ff.

Midterm (see Appendix E)

Topic 10. The General Problem Solver (GPS) program. This was one of the earliest and most influential AI programs (written by a Nobel Prize winner—Herbert Simon), attempting to be able to solve virtually any problem as humans would, using the techniques of means-ends analysis.

Readings: *W, Chap. 5 (to p. 143);
*Slagle, *Artificial Intelligence* (1971), Chap. 8 (contains a detailed, but

elegantly presented, trace of a GPS solution to a puzzle);
Newell and Simon, "Computer Science as Empirical Inquiry," in H;
Dreyfus, *What Computer's Can't Do* (1979): 76ff, 93ff, 112ff.

Topic 11. Sacerdoti's NOAH program—an interactive problem-solving program that is more powerful (more "intelligent") than GPS.

Reading: Sacerdoti, *A Structure* for *Plans and Behavior* (1977): 1-8, 25-27, 55-61 (contains the solution to AIQ question 4), 76-87.

Topic 12. Production Systems and Expert Systems. Trace of the production-system program in W, pp. 357-66. "Expert systems" are among the hottest topics in AI these days, accounting for much of the popularity of AI in industry.

Readings: *W, chap. 5 (pp. 143-56);
*W, Chap. 9;
Pylyshyn, "Complexity and the Study of Artificial and Human Intelligence," in H;
Feigenbaum and McCorduck, *The Fifth Generation* (1983).

Topic 13. Natural-Language Processing.

Readings: *Hendrix and Sacerdoti, "Natural-Language Processing" (1981);
*W, Chap 6.

13A. Weizenbaum's ELIZA program (sometimes called DOCTOR), supplemented by a discussion of the limitations of pattern matching: see Winograd, *Language as a Cognitive Process* (1983): 45-46 and especially exercise 2.2, p. 68.

Readings: *Weizenbaum, "ELIZA" (1966);
Weizenbaum, *Computer Power and Human Reason* (1976), Intro. and Chap. 7.

13B. Kenneth Colby's PARRY program: a simulation of a paranoid mind, using the techniques of ELIZA.

Readings: Colby, "Modeling a Paranoid Mind" (1981);
Weizenbaum, "Automating Psychotherapy" (1974).

Assignment: Write a brief paper (1-2 pages) on Weizenbaum's and Colby's views on ELIZA, PARRY, and the Turing test. (Weizenbaum believes that ELIZA is *not* intelligent but "passes" the test, and is therefore a *reductio* of the test; Colby believes that PARRY has passed the test and is, therefore, a good (partial) model of a mind, albeit a paranoid one.)

13C. Terry Winograd's SHRDLU program: one of the most successful and influential natural-language-processing programs. A (simulated) robot "understands" natural-language commands and is able to plan and execute various actions in a "blocks world."

Readings: *Boden, *Artificial Intelligence and Natural Man* (1977): 114-23;
Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," in H;
McDermott, "Artificial Intelligence Meets Natural Stupidity," in H;
Dreyfus, *What Computers Can't Do* (1979): 5-14.

13D. Roger Schank's Conceptual Dependency theory of knowledge representation. The SAM and PAM programs.

Readings: *Winston, Chap. 7;
Schank and Riesbeck, *Inside Computer Understanding* (1981), Chaps. 1-3,

5-8;
Minsky, "A Framework for Representing Knowledge," in H;
Marr, "Artificial Intelligence," in H;
Dreyfus, *What Computers Can't Do* (1979): 40-47.

Assignment: Hand out the BORIS story and questions (but *not* the answers) given in Appendix F. Students are to read the story and answer the questions, much as they would on a reading comprehension test, and get some friends to do the same. Then give them BORIS's answers. The students should write a brief paper comparing and contrasting the human and computer answers, speculating on possible explanations for any differences.

Topic 14. Further philosophical issues.

14A. Can computers think?

Readings: *Minsky, "Why People Think Computers Can't" (1982);
*McCorduck, "Humans That Think" (1983).

14B What computers can't do.

Readings: *Dreyfus, "From Micro-Worlds to Knowledge Representation," in H;
Marr, "Artificial Intelligence," in H;
McDermott, "Artificial Intelligence Meets Natural Stupidity," in H;
Haugeland, "The Nature and Plausibility of Cognitivism," in H.

14C. Minds, brains, and programs: Searle's Chinese-Room Argument that (roughly) computers will never be able to think because they lack the requisite causal powers.

Readings: *Searle, "Minds, Brains, and Programs" (1980a);
Abelson, *et al.*, "Open Peer Commentary" (1980);
Searle, "Author's Response" (1980b);
Searle, "The Myth of the Computer" (1982);
Rapaport, "Machine Understanding and Data Abstraction in Searle's Chinese Room" (1985);
Rapaport, "Searle's Experiments with Thought" (1986a);
Rapaport, "Philosophy, Artificial Intelligence, and the Chinese Room Argument" (1986b).

Appendix A

John Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence* (Cambridge, Mass.: MIT/Bradford, 1981).

Annotated Table of Contents

John Haugeland, "Semantic Engines: An Introduction to Mind Design." Excellent introduction to the fields of AI and cognitive science.

Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search." Good background reading for the General Problem Solver (GPS) program.

Zenon Pylyshyn, "Complexity and the Study of Artificial and Human Intelligence." Discusses production systems and the relation of AI to psychology.

Marvin Minsky, "A Framework for Representing Knowledge." A classic paper in AI; discusses frames and the neat/scruffy debate on the role of logic in AI.

David Marr, "Artificial Intelligence—A Personal View." Interesting remarks on natural language, against Dreyfus; discusses production systems, scripts.

Drew McDermott, "Artificial Intelligence Meets Natural Stupidity" Discusses semantics of data structures used in AI; should be read *after* Dreyfus's essay.

Hubert L. Dreyfus, "From Micro-Worlds to Knowledge Representation: AI at an Impasse." A classic argument against the possibility of AI succeeding; excerpted from his *What Computers Can't Do* (1979).

Hilary Putnam, "Reductionism and the Nature of Psychology." Good defense of functionalism.

Daniel C. Dennett, "Intentional Systems." Excellent discussion of functionalism, modular design, and top-down design and stepwise refinement (two techniques typically taught in introductory computer science courses); contrast with McDermott's and Searle's essays; should be read in conjunction with Putnam's essay.

John Haugeland, "The Nature and Plausibility of Cognitivism." Compare with Dennett's essay; Sect. 7 should be compared with Dreyfus's essay.

John R. Searle, "Minds, Brains, and Programs." A modern classic; good on SAM, Turing test; compare McDermott's essay. The full essay, together with objections by other cognitive scientists and replies by Searle, is in *Behavioral and Brain Sciences* 3 (1980) 417-57; follow-ups are Searle's review of Hofstadter and Dennett's *The Mind's I* (1981): "The Myth of the Computer," *New York Review of Books* (29 April 1982) and Searle's *Minds, Brains and Science* (1984).

Jerry A. Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." Good background for SHRDLU (Winograd's natural-language-understanding program that is a practicing solipsist—though it doesn't know it); compare McDermott's essay.

Donald Davidson, "The Material Mind." Discusses a robot that can pass the Turing test because it exactly simulates a human.

Appendix B: Suggested Term Projects

Suggested Paper Topics:

(1) A description and critique of any AI program or an area of research not covered in class.
(2) A philosophical analysis of any essay(s) from Haugeland 1981.
(3) Discussion of Dreyfus's critique of any program covered in class.
(4) Discussion (book review) of Feigenbaum and McCorduck's *The Fifth Generation* (1983).
(5) Expansion of any Midterm question or any previous written assignment.
(6) Discussion of the "neat/scruffy" debate between those AI researchers who hold logic to be central and those who disdain its use.
(7) Discussion of Minsky's "Jokes and the Logic of the Cognitive Unconscious" (1980). [This was the topic of Minsky's Academic Convocation lecture at Fredonia the semester the course was given.]

Suggested Programming Projects:

(1) Implement the production system in Winston 1977, Chap. 15.
(2) Implement any of the programs in Schank and Riesbeck 1981.
(3)Implement SIR (a question-answering system) as given in Shapiro 1979.
(4) Implement ELIZA in LISP.
(5) Translate ELIZA into French, German, or Spanish.
(6) Write a working editor for LISP. (See Charniak *et al.* 1980, Chap. 7.)
(7) Learn LISP, using Shapiro (1986).

Notes: Projects 1-3 required copying and debugging programs. Once implemented, they can be modified as future exercises or projects; see the appropriate texts. Similarly, project 4 was a copy-and-debug project; we had a listing of this program, but had been having trouble translating it into the dialect of LISP available on our computer. No one tried project 5 at Fredonia, but I have since had a student at Buffalo who translated it into French; it is a worthy term project with possible beneficial side effects as a tool for foreign-language departments. Project 6 calls for great LISP skill; our system did not have a working editor.

Appendix C

Artificial IQ Test

1. In each of the following, A is to B as C is to? (Illustration on page 112 is from Winston, 1977: 27.)
2. Describe how you solved No. 1(b).
3. On the basis of the following two computer/human interactions, which program exhibits more intelligence? (COMPUTER OUTPUT IS IN CAPS; human input in lower case.)

Program "B":

PROCESSING PARAGRAPH 1:

Richard had not heard from his college roommate Paul for years. Richard had borrowed money from Paul which was never paid back. But now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was.

READY TO ACCEPT Q/A:

What happened to Richard at home? RICHARD GOT A LETTER FROM PAUL.
Who is Paul? RICHARD'S FRIEND.
Did Richard want to see Paul? YES. RICHARD WANTED TO KNOW HOW PAUL WAS.
Had Paul helped Richard? YES. PAUL LENT MONEY TO RICHARD.
Why did not Richard pay Paul back? RICHARD DID NOT KNOW WHERE PAUL WAS.
How did Richard feel when the letter appeared? RICHARD FELT GLAD BECAUSE PAUL AND HE WERE FRIENDS.

PROCESSING PARAGRAPH 2:

Unfortunately, the news was not good. Paul's wife Sarah wanted a divorce. She also wanted the car, the house, the children, and alimony. Paul wanted the divorce, but he did not want to see Sarah walk off with everything he had. His salary from the state school system was very small. Not knowing who to turn to, he was hoping for a favor from the only lawyer he knew. Paul gave his home phone number in case Richard felt he could help.

READY TO ACCEPT Q/A:

What was the letter about? PAUL AND SARAH WERE GETTING A DIVORCE.
Did Paul write Sarah? NO. IT WAS NOT SARAH. IT WAS RICHARD WHO GOT THE LETTER.
Why was Paul upset about the divorce? PAUL AND SARAH WERE FIGHTING OVER THE FAMILY POSSESSIONS.

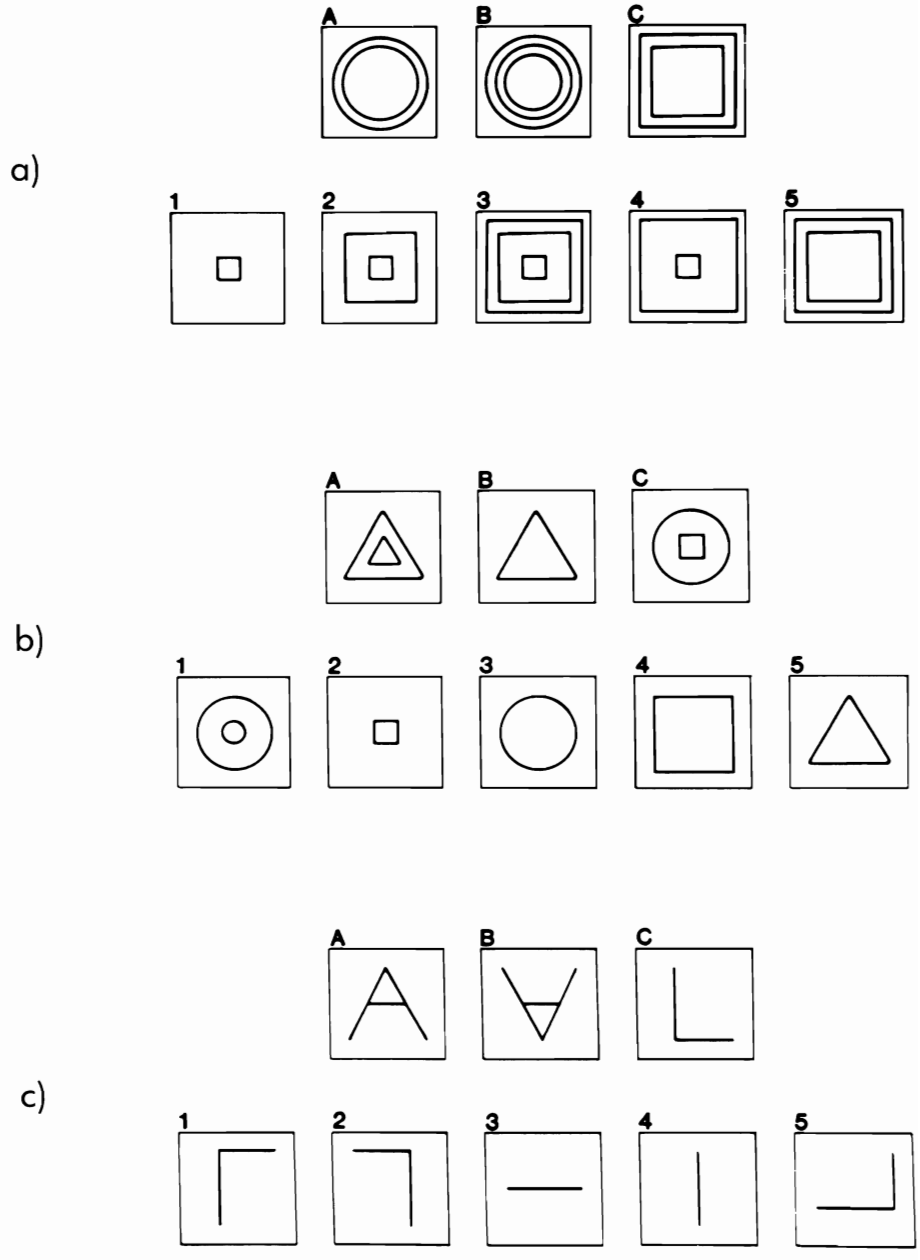


Figure 1. Geometric analogies.
(From Winston 1977:27)

What did Paul do for a living? PAUL WAS A TEACHER.
What did Sarah want? SARAH WANTED TO HAVE THE CAR AND THE HOUSE AND THE KIDS AND THE ALIMONY.
Why did Paul write to Richard? PAUL WANTED RICHARD TO BE HIS LAWYER.

(From Lehnert *et al.* 1983: 16-17.)

Program "D":

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
It's true. I am unhappy.
DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY
I need some help, that much seems certain.
WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP
Perhaps I could learn to get along with my mother.
TELL ME MORE ABOUT YOUR FAMILY
My mother takes care of me.
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU
My father.
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don't want me to notice that.
WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE
You don't argue with me.
WHY DO YOU THINK I DON'T ARGUE WITH YOU
You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU
My father is afraid of everybody.
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

(From Weizenbaum 1966: 36-37. Note: For best results, do *not* include these acknowledgments in a classroom version!)

4. Solve the following problem. (HINT: The solution requires 21 steps.) (From Sacerdoti 1977: 55-56.)

In a room there is a robot with no sensory ability whatsoever. Four places are defined in the room: BOX1, BOX2, TABLE, and DOOR. Outside the room there is one place: OUTSIDE.

At DOOR there is a non-empty pile of red objects.

At BOX 1 or BOX 2 (we don't know which) there is a non-empty pile of keys, all of which fit the door. We don't know what is at the other box.

TABLE is empty.

The robot has three possible actions:

1. Pick up—If the robot is holding something, this action has no effect. Otherwise, some object at the location will be in the robot's hand when this action is completed.
2. Put down—If the robot is not holding anything, this action has no effect. Otherwise, the object in the robot's hand is added to the pile at the current location of the robot.
3. Go to X—The robot's location becomes X. If X is OUTSIDE, there must be a key in the pile at DOOR or this action has no effect.

The robot has no way to tell if any action had an effect or not.

Initially the robot is at an undetermined place in the room. Figure 2 suggests the initial configuration. The problem is to develop a set of actions that will ensure that a red object is OUTSIDE.

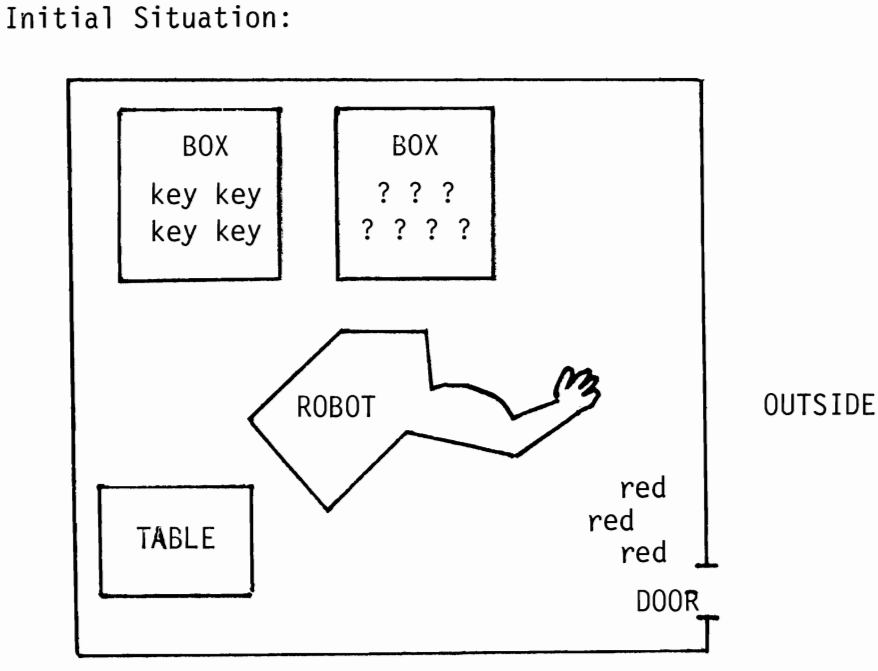
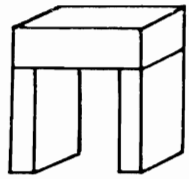


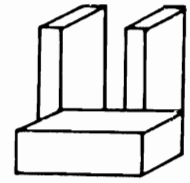
Figure 2. The "Keys and Boxes" Problem (From Sacerdoti 1977)

5. (Illustrations from Winston 1977:29.)

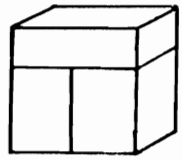
A. This is a CRDL:



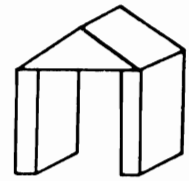
B. This is not a CRDL:



C. This is not a CRDL:



D. This is a CRDL:



(a) Define CRDL.
(b) How did you arrive at your definition?

6. What does the following sentence mean? (You may parse it, or draw a picture, or paraphrase it.)

JOHN PUT THE HAMMER IN THE TOOLBOX ON THE TABLE.

7. Consider the following story:

John decided to visit Bill. He drove to his house, but he saw that the lights were off, so he went to the movies.

a) Did John see Bill?
b) Whose house was dark?
c) Who went to the movies?
d) How did you determine the answers to these questions?

8. Consider the following lead paragraph from a recent news item:

Twenty-five women passed the strenuous New York City Fire Department test of physical endurance today. This was the first test given since women were allowed to join the Department.

What topics were covered in the rest of the article?

9. Consider the following story:

John went out to a restaurant last night. He ordered a steak. When he paid for it, he noticed that he was running out of money. He hurried home since it had started to rain.

Did John eat dinner last night? How do you know? How would a computer know?

10. Consider the following story:

Willa was hungry. She reached for the phone book. She got in her car.

Did Willa eat the phone book? How do you know? How would a computer know?

Appendix D

Some Definitions of Artificial Intelligence

1. The goal of work in artificial intelligence is to build machines that perform tasks normally requiring human intelligence. (Nilsson 1971: vii.)
2. Research scientists in Artificial Intelligence try to get machines to exhibit behavior that we call intelligent behavior when we observe it in human beings. (Slagle 1971: 1.)
3. B. Raphael...has suggested that AI is a collective name for problems which we do not yet know how to solve properly by computer. (Michie 1971: 101.) [Note that it follows that once we do know how to solve them, they are no longer AI!]
4. What is or should be [AI researchers'] main scientific activity—*studying the structure of information and the structure of problem solving processes independently of applications and independently of its realization in animals or humans*. (McCarthy 1974: 317.)
5. By "artificial intelligence" I therefore mean the use of computer programs and programming techniques to cast light on the principles of intelligence in general and human thought in particular. (Boden 1977: 5.)

Note: A valuable discussion of the nature of AI may be found in the "debate" between Schank (1983) and Bundy (1983).

Appendix E

Midterm Examination

Write 1-2 pages each on 3 of the following questions:

1. Describe Functionalism as a theory designed to solve the mind-body problem. How does it differ from the Identity Theory?
2. Describe the ANALOGY program. Does it exhibit intelligence?
3. How would Turing respond to the objection that a program might pass the Turing test, yet still not be able to think?
4. Suppose that Winston's learning program is taught what a house is. [Note: This was done in class, using simple line drawings and Winston's semantic network technique as described in his text.] Has it really learned what a house is? Has it learned anything?

Appendix F

Assignment for Topic 14 D

READ THE FOLLOWING PASSAGE AND ANSWER THE QUESTIONS. (*Note*: The version handed out in class should omit BORIS's answers, which are given here in small CAPS.)

John left his office early Friday afternoon so he could attend a going-away party for his friend Bill. Bill had been transferred to California. When he arrived there were lots of people and plenty of beer. Three years ago Bill had borrowed $50 from John which John hoped to get back before Bill left. He was about to remind Bill about it when he was called to the phone. It was his wife, Mary, and she was hysterical. Mary had just received a call from a stranger who claimed to have kidnapped their daughter Susan. He said he would call again. John told Mary to stay at home, keep off the phone, and wait until he got there.

John drove home immediately. Just as he walked in the door, the phone rang. John picked it up. But it was only Karen, one of Mary's friends. John told her Mary was out and hung up. Mary wanted to call the police but John thought they should wait. They were discussing this when Susan walked in the door. She had stayed after school for a gymnastics meet and taken a late bus home. Nothing had happened to her. It was all a hoax.

1. Why did John go to the party? HE WANTED TO HAVE A GOOD TIME AND HE WAS BILL'S FRIEND AND HE WANTED TO GET 50 DOLLARS BACK FROM HIM
2. Why did John leave the party? HE WAS WORRIED ABOUT SUSAN
3. Why? SUSAN WAS KIDNAPPED
4. What happened to John at the party? HE GOT A CALL FROM MARY
5. Why did Mary call John at the party? SHE WANTED TO TELL HIM ABOUT SUSAN'S KIDNAPPING
6. Why did John want money from Bill? HE BORROWED 50 DOLLARS FROM HIM
7. Why didn't John talk to Bill? HE WAS MORE WORRIED ABOUT SUSAN
8. Who called John at home? KAREN
9. Why? SHE WANTED TO TALK TO MARY
10. Was Karen upset? NO. SHE DID NOT KNOW ABOUT SUSAN'S KIDNAPPING
11. Did John kidnap Susan? NO. HE WAS HER FATHER
12. Where was Susan? SHE WAS ATTENDING A GYMNASTICS MEET
13. What happened when Susan came home? JOHN WAS RELIEVED THAT SHE WAS SAFE

(From Dyer 1982: 287-88.)

Notes

The preparation of this article was supported in part by SUNY Buffalo Research Development Fund grant #150-9216-F. I am grateful to the SUNY Fredonia Department of Philosophy for a leave of absence during the 1982-83 academic year; to Shoshana Hardt, for many of the ideas for this course; and to Stuart C. Shapiro, for discussions on many of these issues.

1. Since this article was written, I have taught AI using Charniak and McDermott 1985, and found it unsatisfactory.

2. A new book by John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT Press, 1985), would also make an excellent supplementary text.

References

Abelson, Robert P., *et al.*, "Open Peer Commentary [on Searle 1980a]," *Behavioral and Brain Sciences* 3 (1980) 424-50.

Anderson, A. R., ed., *Minds and Machines* (Englewood Cliffs, N. J.: Prentice-Hall, 1964).

Barr, Avron, Edward A. Feigenbaum, and Paul R. Cohen, eds., *The Handbook of Artificial Intelligence*, 3 volumes (Los Altos, CA: William Kaufmann, 1981-1982).

Boden, Margaret, *Artificial Intelligence and Natural Man* (New York: Basic Books, 1977).

Bundy, Alan, "The Nature of AI: A Reply to Schank," *AI Magazine* 4.4 (Winter 1983) 29-31.

Charniak, Eugene, and Drew McDermott, *Introduction to Artificial Intelligence* (Reading, MA: Addison-Wesley, 1985).

_____, C. K. Riesbeck, and D. V. McDermott, *Artificial Intelligence Programming* (Hillsdale, N. J.: Lawrence Erlbaum Associates, 1980).

Churchland, Paul M., *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (Cambridge, MA: MIT/Bradford, 1984).

Cohen, Paul R., and Edward A. Feigenbaum, eds., *The Handbook of Artificial Intelligence*, Vol. III (Los Altos, CA: William Kaufmann, 1982).

Colby, Kenneth Mark, "Modeling a Paranoid Mind," *Behavioral and Brain Sciences* 4 (1981) 515-60.

Davis, Martin, *Computability and Unsolvability* (New York: McGraw-Hill, 1958).

Dennett, Daniel C., "Artificial Intelligence as Philosophy and as Psychology," in Ringle (1979): 57-77; reprinted in D. C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, MA: MIT/Bradford, 1978): 109-26.

_____, "Reflections [on Hofstadter 1981]," in Hofstadter and Dennett 1981: 92-95.

Dreyfus, Hubert L., *What Computers Can't Do: The Limits of Artificial Intelligence* (New York: Harper and Row, rev. ed., 1979).

Dyer, Michael G., *In-Depth Understanding*, Research Report No. 219 (Yale University Dept. of Computer Science, 1982).

Evans, Thomas G., "A Program for the Solution of Geometric-Analogy Intelligence Test Questions," in M. Minsky, ed., *Semantic Information Processing* (Cambridge, MA: MIT Press, 1968).

Feigenbaum, Edward A., and Pamela McCorduck, *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World* (Reading, MA: Addison-Wesley, 1983).

Fodor, Jerry, "The Mind-Body Problem," *Scientific American* 244 (Jan. 1981) 114-23.

Gardner, Howard, *The Mind's New Science: A History of the Cognitive Revolution* (New York: Basic Books, 1985).

Friedman, Daniel P., *The Little LISPer* (Chicago: Science Research Associates, 1974).

Haugeland, John, ed., *Mind Design: Philosophy, Psychology, Artificial Intelligence* (Cambridge, MA: MIT Press, 1981).

Hendrix, Gary G., and Earl D. Sacerdoti, "Natural-Language Processing," *Byte* 6 (September 1981) 304-52.

Hofstadter, Douglas R., "The Turing Test: A Coffeehouse Conversation," in Hofstadter and Dennett 1981: 69-92.

_____, "The Pleasures of Lisp: The Chosen Language of Artificial Intelligence," *Scientific American* 248 (Feb. 1983a) 14-28; reprinted as "Lisp: Atoms and Lists," in Hofstadter 1985: 396-409.

_____, "Tripping the Light Recursive in Lisp, the Language of Artificial Intelligence," *Scientific American* 248 (Mar. 1983b) 22-31; reprinted as "Lisp: Lists and Recursion," in Hofstadter 1985: 410-24.

_____, "In which a Discourse on the Language Lisp Concludes with a Gargan-

tuan Italian Feast," *Scientific American* 248 (April 1983c) 14-28; reprinted with additions as "Lisp: Recursion and Generality," in Hofstadter 1985: 425-54.

_____, "Mind, Body and Machine," *New York Times Book Review* (13 Nov. 1983d).

_____, *Metamagical Themas: Questing for the Essence of Mind and Pattern* (New York: Basic Books, 1985).

_____ and Daniel C. Dennett, eds., *The Mind's I* (New York: Basic Books, 1981).

Lehnert, W. G., M. G. Dyer, P. N. Johnson, C. J. Yang, and S. Harley, "BORIS—An Experiment in In-Depth Understanding of Narratives," *Artificial Intelligence* 20 (1983) 15-62.

McCarthy, John, Review of "Artificial Intelligence: A General Survey," *Artificial Intelligence* 5 (1974) 317-22.

McCorduck, Pamela, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (San Francisco: W. H. Freeman, 1979).

_____, "Humans That Think," *AI Magazine* 4.3 (Fall 1983) 35.

Michie, Donald, "Formation and Execution of Plans by Machine," in N.V. Findler and B. Meltzer, eds., *Artificial Intelligence and Heuristic Programming* (New York: American Elsevier, 1971): 101-24.

Minsky, Marvin, "Jokes and the Logic of the Cognitive Unconscious," Memo No. 603 (MIT Artificial Intelligence Laboratory, 1980).

_____, "Why People Think Computers Can't," *AI Magazine* 3:4 (Fall 1982) 3-15.

Moulton, Janice, and Voytek, Jane, "Can Humans Think Machines Think?" *Teaching Philosophy*, 3:2 (1979) 153-67.

Nilsson, Nils J., *Problem-Solving Methods in Artificial Intelligence* (New York: McGraw-Hill, 1971).

Putnam, Hilary, "Minds and Machines" (1960), in Anderson 1964: 72-97.

Rapaport, William J., "Critical Thinking and Cognitive Development," *APA Newsletter on Pre-College Instruction in Philosophy* 1 (Spring/Summer 1984a) 4-5; reprinted in *Proc. American Philosophical Assoc.* 57 (1984b) 610-15.

_____, "Searle's Experiments with Thought," *Philosophy of Science* (1986a); preprinted as Tech. Report No. 216 (SUNY Buffalo Dept. of Computer Science, 1984c).

_____, "Machine Understanding and Data Abstraction in Searle's Chinese Room," *Proc. Cognitive Science Society* 7 (1985) 341-45.

_____, "Philosophy, Artificial Intelligence, and the Chinese Room Argument," *Abacus* (Summer 1986b).

Rich, Elaine, *Artificial Intelligence* (New York: McGraw-Hill, 1983).

Ringle, Martin D., ed., *Philosophical Perspectives in Artificial Intelligence* (Sussex, Eng.: Harvester, 1979).

Sacerdoti, Earl, *A Structure for Plans and Behavior* (Amsterdam: Elsevier, 1977).

Schagrin, Morton L., William J. Rapaport, and Randall R. Dipert, *Logic: A Computer Approach* (New York: McGraw-Hill, 1985).

Schank, Roger C., "The Current State of AI: One Man's Opinion," *AI Magazine* 4.1 (Winter-Spring 1983) 3-8.

_____, and Christopher K. Riesbeck, eds., *Inside Computer Understanding:*

Five Programs Plus Miniatures (Hillsdale, NJ: Lawrence Erlbaum, 1981).

Searle, John R., "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3 (1980a) 417-24; reprinted in Haugeland 1981: 282-306, and in Hofstadter and Dennett 1981: 353-73.

_____, "Author's Response: Intrinsic Intentionality," *Behavioral and Brain Sciences* 3 (1980b) 450-57.

_____, "The Myth of the Computer," *New York Review of Books* (29 April 1982) 3-6.

_____, *Minds, Brains and Science* (Cambridge, MA: Harvard University Press, 1984).

Shapiro, Stuart C., *Techniques of Artificial Intelligence* (New York: D. Van Nostrand, 1979).

_____, *LISP: An Interactive Approach* (Rockville, Md.: Computer Science Press, (1986).

_____, ed., *Encyclopedia of Artificial Intelligence* (New York: John Wiley, forthcoming).

Slagle, James R., *Artificial Intelligence: The Heuristic Programming Approach* (New York: McGraw-Hill, 1971).

Taylor, Richard, *Metaphysics* (Englewood Cliffs, NJ: Prentice-Hall, 3rd ed., 1983).

Touretzky, David S., *LISP: A Gentle Introduction to Symbolic Computation* (New York: Harper and Row, 1984).

Turing, A. M., "Computing Machinery and Intelligence," *Mind* 59 (1950). Reprinted in Anderson 1964: 4-30, and excerpted in Hofstadter and Dennett 1981: 53-67.

Wand, Mitchell, "What is LISP?", *American Mathematical Monthly* 91 (1984) 32-42.

Webber, B. L., and N. J. Nilsson, eds., *Readings in Artificial Intelligence* (Palo Alto: Tioga, 1981).

Weizenbaum, Joseph. "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the Association for Computing Machinery* 9 (1966) 36-45. Reprinted in *CACM* 26 (1983) 23-28.

_____, "Automating Psychotherapy," *ACM Forum* 17 (1974) 543. Reprinted in *CACM* 26 (1983) 28.

_____, *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman, 1976).

Wilensky, Robert, *LISPcraft* (New York: W. W. Norton, 1984).

Winograd, Terry, *Language as a Cognitive Process*, Vol. 1: *Syntax* (Reading, MA: Addison-Wesley, 1983).

Winston, Patrick Henry, *Artificial Intelligence* (Reading, MA: Addison-Wesley, 1st ed., 1977).

_____, *Artificial Intelligence* (Reading, MA: Addison-Wesley, 2nd ed., 1984).

_____, and B. K. P. Horn, *LISP* (Reading, MA: Addison-Wesley, 2nd ed., 1984).

William J. Rapaport, Computer Science, State University of New York, Buffalo, New York 14260 USA