Book Review
for
Minds and Machines

Francis Jeffry Pelletier
Department of Computing Science
Department of Philosophy
University of Alberta
Edmonton, Alberta
Canada T6G 2E1

Todd C. Moody *Philosophy and Artificial Intelligence* (Englewood Cliffs, NJ: Prentice Hall). viii+ 175pp.

I was asked to develop a course “Philosophy and Cognitive Science” to be taught for the first time in Spring 1995 in the Philosophy Department at the University of Alberta. Since my cognitive science-related interests are focussed more towards philosophy mixed with artificial intelligence (AI) and linguistics than towards (say) neuroscience or anthropology, I decided to slant the course in that direction. The departmental intent was that this should be an upper-level course, but with no specific prerequisite courses. This meant that while there was a “three previous courses in philosophy” prerequisite, the students could not be expected to have taken any particular course, as (say) a philosophy of mind course or a logic course. Further, I had in mind that the course would be part of an initiative to create an undergraduate program in Cognitive Science at the University of Alberta, and so I encouraged students from linguistics, computer science, and psychology who had taken courses in their department like “Language and Mind”, “Introduction to AI”, and “Introduction to Cognitive Science” to take this course even without the philosophy background. This resulted in the class consisting of about a quarter each philosophy, psychology and computer science majors, with the rest being drawn from linguistics, education, and anthropology. And although this distribution of students arguably lessened the philosophical sophistication of both the lectures and the class discussion, I (and the students, I believe) found the perspectives brought to the topic by these other students to be fascinating and quite provocative.

Because of my decision to slant the course toward AI, I sought a book that would explain a sufficient amount of AI background (aimed at my non-computer science students), would discuss the role of language in intelligence and explain some of the works in natural language understanding, and (of course) would emphasize such philosophico-cognitive science issues as “Can Computers Think?”. My thoughts about appropriate topics was also heavily influenced by William Rapaport’s article (1986), although I hoped to bring it up to date (Rapaport’s article described a course taught in 1983).

There are many books one could choose for such a course. Here are some: Boden (1987, 1988), Born (1987), Broadbent (1993), Crockett (1994), Fetzer (1991), Garnham (1987), Haugeland (1985), McClintock (1995), Moody (1993), Morelli *et al* (1992), Robinson (1992). None of these books really meet all my desires in a textbook for this class; especially I found that they were all weak on the topic of natural language understanding (with the possible exception of Boden, and the article by Gazdar in Broadbent’s book). For example, Garnham is mostly descriptive of AI and only the second-last chapter concerns “conceptual issues” while Broadbent is an anthology in which a number of well-known British researchers describe how they see the relationship between cognitive science and his/her research area. And although I appreciate that opposition to the very possibility of AI and Cognitive Science is still a legitimate part of Philosophy of AI and Cognitive Science, I still think it best not to use a work that is blatantly opposed to the whole field as a textbook, and so I would rule out Born, Crockett and McClintock. The books remaining on my list are all suitable for some course on the topic (although one might think that Boden 1987 is a bit dated, even with all the update material in this second edition); a teacher’s adopting one or another of them would be a decision based on the level of the course, a preference for the author’s specific emphasis on certain topics, and the writing style.

I chose Moody’s book. Some of the books on my list above came to my attention after I had adopted his as my textbook, but I probably would have chosen Moody’s anyway because of its plain, yet very pleasant and engaging style, its (pretty much) error-free presentation, and its reasonable cost. One drawback to it is that it seems to be written at too elementary a level in its discussion of phil osophical matters (as opposed to its discussion of factual AI content, which, when present, was also very elementary but perhaps was at a more appropriate level for this course) for upper level philosophical students...or even for the students in my “mixed” class. With an eye to remedying this, I also ordered Fitzer’s book, which concerns many of the same philosophical matters at a more sophisticated level. This was only partially successful, as the majority of my non-philosophy students found it “obsessively obscure”, while the philosophy students objected to (what they perceived as) its “phenomenological bias.” (Both comments appeared on anonymous class evaluations).

Moody makes two points clear both in the preface and at various places throughout the book: that this is a *philosophical* book and not a technical account of AI, and that it is *introductory* and not written just for philosophers or philosophy students. He is right on both counts, bringing along both the benefits and the shortcomings of this genre. As I have said, those students from outside philosophy generally found the book excellent in its introductory coverage. (I like to think that they benefited also from my lectures that were pitched at a somewhat higher level!) Exceptions to this judgment came not only from the philosophy students but also from the computer science students, who found the description of “Computing Machines” (Chapter 3) to be so elementary as to miss the area of modern computing programs and machines completely. This chapter introduces the notion of a formal system, distinguishing syntax from semantics in a very rough-and-ready way, discusses elementary logic at the truth-table level, discusses the notion of rationality, introduces Turing machines and von Neumann machines, discusses what it means to be a “digital device” and to “follow a rule”, and tries to say what high-level programming languages are. That all this is carried out in 25 (small) pages of text should give some idea of the level of detail at which Moody explains matters, and should give some force to the computer students’ complaints. Of course, all these topics are crucial for the course content, and an upper level course should expect the students to be able to follow these topics at about the level of Haugeland (1985) Chapters 2 and 3. So if the teacher of such a course adopts Moody’s book, s/he will have to be prepared to give much other material in lecture.

My students, even the computing ones (who seemed to be taught none of it in Computer Science), wanted more about the history of computing (and of AI) than Moody offers in his Chapter 1. I found the summaries in Boden (1987, 1988) to be the sort of thing the students were looking for. Another area that Moody does not discuss concerns the mathematical challenges to AI based on Gödel's theorems (and popularized by Lucas 1961 and Penrose 1989, 1994). Many of my students had heard some version of these arguments and wanted to know more. The discussion in Robinson (1992) Chapter 6 is quite good, especially in its explaining at a level appropriate to my students.

The other background topic discussed by Moody concerns the mind-body problem. Chapter 2 gives a pretty reasonable introduction to the entire topic, motivating dualism and then discussing various reactions to dualism. Moody mentions the identity theory, although he does not choose to go into the various flavors of it; and he discusses behaviorism, distinguishing methodological from philosophical versions (attributing the latter to the logical positivists, whose doctrines on meaning are briefly discussed). Finally, functionalism is characterized and said to be "the favorite theory of mind among AI-advocates." If a course in Philosophy of AI and Cognitive Science were to have a prerequisite of a philosophy of mind course, this introduction to the topic would be 'way too elementary. In my mixed background class, only the philosophy majors objected to Moody's skipping over various of the ins and outs of theories of mind and to his omission of epiphenomenalism (which some students think is the only reasonable way to combine functionalism with the facts about qualia!).

After the background topics, the main theme of any course in Philosophy of AI is introduced: The Turing Test and "strong AI". This short chapter 4 also brings forward for consideration Block's "jukebox" (in conjunction with ELIZA) as intuitions opposed to the Turing Test. And the chapter then moves on to the Chinese Room argument, which is described pretty well (although the various "responses" to it are not mentioned--except for the Systems Reply (which is brushed aside) and the Robot Reply (which Moody supports at some length). These topics have become part of the very center of Cognitive Science and Philosophical AI, and an instructor will certainly want to make more of them than what is presented in the text, although what is presented is well enough done.

Chapter 5, "The Nature of Intelligence", is the main chapter of the book, at least in terms of the importance of the topics introduced. Moody starts by contrasting chess-playing with face-recognition, asking which requires more intelligence. This in turn gives rise to discussions of the difference between algorithms and heuristics, to tacit vs. explicit knowledge, and to whether recognition, generally speaking, is perhaps a very broad ability that is applied to a variety of tasks or whether it is instead maybe "an aspect of understanding itself [and is therefore] interdependent with our ability to make sense [of things we recognize]" and "these faculties [would] in turn only make sense against a backdrop of interests, commitments, desires, fears," etc. Since such a view of human recognition is widely different from the "pattern-matching" paradigm employed in (most) computational systems, Moody wonders whether it matters *how* an artificial system accomplishes its recognitions, or just that it succeeds in recognizing something. "If we regard the mind simply as a collection of task-oriented modules, then clearly nothing important hinges on mimicking the specific actions of the human brain....If on the other hand we have an understanding of the mind according to which mental and biological characteristics interpenetrate and are interdependent then clearly it will not be sufficient to stray too far from the concretely human details of the mental." Moody's sympathies lie with the latter alternative.

This chapter also discusses language, certainly a central issue when you set the Turing Test as the focal point of the course. Moody has a pretty reasonable discussion of philosophical issues concerning communication generally. The discussion of linguistics and computational linguistics (or natural language understanding systems) seems to me to be much less successful. Moody claims that it is impossible to parse without semantic information; and while this is possibly so, I do not think it ought to be presented as obviously true. And when considering the pair of sentences "Time flies like an arrow. Fruit flies like a banana", Moody says that a fluent speaker can instantly understand these two sentences without difficulty, but that writing an algorithm capable of discriminating the structural difference "is something else again." He seems to think that the surface structure of the two sentences is "closely parallel" even though "it is obvious that the same rules cannot be applied to the two sentences." Most linguists and computational linguists would say that the surface structures of the two sentences are widely different. Moody also claims that examples such as these "have led ... Chomsky to suppose that ...there must be a 'deep structure' underlying these surface structures that conforms to the principles of a formal system." And from there we jump to the claim that

Chomsky believes this underlying formal system to be "the universal grammar." Moody has his doubts as to whether there is any such universal grammar ("neither Chomsky nor his followers have had a great deal of success in devising a theory of the universal grammar"); and he has doubts as to whether even if it *is* constructed that it will be "of a form that can be represented in a purely formal system". In conjunction with this, it is claimed that computational linguistics is "the research program that is based on the hypothesis that natural languages can be shown to be formal systems", and apparently if there is no formal universal grammar Moody thinks computational linguistics must fail. Maybe, he says, "our ability to use language is an outgrowth and extension of other more or less biological abilities and involvements with the world", and he apparently thinks that this goes against computational linguistics. But attempting to be fair to everyone, Moody admits "my skepticism is as conjectural as the computational linguist's optimism" about the universal grammar being a formal system. My experience in computational linguistics is that many (most?) either do not believe that there is "a universal grammar" of the sort Moody is discussing or else believe that it does not impact in any way on their research.

The remainder of this chapter discusses consciousness and self-awareness, tying this in with subjectivity and objectivity (what's it like to be a bat?) and with issues concerning qualia (of the inverted spectrum type). I would have liked it had Moody also talked about missing qualia and about future neuroscientists who are raised in black-and-white environments. The chapter closes with a discussion of the question of whether something could be intelligent without being conscious. Moody's "idea is that while consciousness may be intimately associated with those states we call mental states, it is not just another one of those mental states....consciousness itself remains distinguishable from them."

Chapter 6 is called "Connections", leading one to think it is gong to be about connectionism. But although this may be the intent of the chapter, the fact is that there is so little factual information about connectionist networks that no student will be able to understand what they do, even if the student already does know what traditional computational models are. The chapter opens with a brief discussion of "symbolic processing", and challenges it with the "problem of knowing language" (which is alleged to be a matter both of sentence processing and of understanding what things to talk about, and that these aspects are "interpenetrating" and presuppose some sort of holism). This leads to a discussion of rule-following, which could have been quite exciting but Moody instead decides to keep this at the level of a "coded" vs. "emergent" distinction and use this distinction to introduce the notion of connectionism – in which Moody claims all rules are emergent. The chapter closes with an interesting discussion about grounds for attributing causal powers to computational devices, and Moody claims that connectionism evades the Chinese Room argument. (In this regard Moody claims that "a connection machine is not just a digital computer 'under a different description'...at least, there is nothing to be gained by so characterizing it"; but he does not consider Fodor & Pylyshyn's (1988) claim that any interesting computation that is performed by a connection machine is "a mere implementation" of a symbolic architecture.)

Chapter 7, the last chapter, is a more broad-ranging discussion of other philosophical issues related to AI, such as personal identity and some ethical issues that arise from the possibility of having intelligent mechanization of activities that now require humans. Moody considers automated psychoanalysis (and cites Weizenbaum's 1976 worries about allowing it), and expert systems in general (and cites worries that children will eventually identify themselves as a type of machine). There are other topics I would have liked to have seen discussed in this chapter, such as whether machines should be granted rights and privileges now reserved for humans, if AI ever succeeds in producing an intelligent artefact. And whether, if it should be shown that this goal of AI is even *possible*, perhaps we should re-think the ethical background of even attempting to carry it out. After all, we now believe that manipulations of human genes can lead to the creation of certain types of people, and we convene expert panels to investigate our ethical qualms concerning whether this should be allowed to proceed. If the goal of AI is even just possible, then perhaps we should take the same attitude toward it and institute panels to look at the possible ethical issues that are inherent in the construction of artificial intelligence.

Despite the criticisms I have levelled at certain aspects of Moody's book, I think it is pretty good. The problem is that there seems to be no natural classroom market for it. I have mentioned that

, by the author's design, it is very elementary–as elementary as any Philosophy 101 textbook (and considerably shorter than most). Yet, it seems very unlikely that "philosophy of AI" will be the introductory class at any school. The natural place for a "philosophy of AI" course is at the upper level, after students have taken the usual introduction to, history of, and central topics in philosophy. But by this point in a student's career Moody's book will be too elementary, especially in its account of philosophy of mind, philosophy of language, and epistemology. Maybe the best market for the book is not as a textbook at all but rather for "educated laypeople".

A different shortcoming, one that remains even if the philosophical level of the course were perfect, is that there is not enough factual information about what sort of things AI programs can actually accomplish. After all, if the Turing Test is the central topic of the course, then, since this is a behavioral test, we would wish to see what sorts of behavior AI programs can evince. Rapaport's (1986) course outline presented a number of famous accomplishments of different AI programs, from analogies to language understanding to problem solving. But these examples were mostly from the 1960's. We hope that the field of AI has advanced since that point.

Bibliography

Boden, Margaret (1987) *Artificial Intelligence and Natural Man* 2nd expanded edition (Cambridge MA: MIT). (1st edition 1977, Harvester Press).

Boden, Margaret (1988) *Computer Models of Mind* (Cambridge: Cambridge UP).

Born, Rainer (1987) *Artificial Intelligence: The Case Against* (NY: Croom Helm).

Broadbent, Donald (1993) *The Simulation of Human Intelligence* (Cambridge, MA: Blackwells).

Crockett, Larry (1994) *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence* (Norwood NJ: Ablex).

Fetzer, James (1991) *Philosophy and Cognitive Science* (NY: Paragon House).

Fodor, Jerry & Zenon Pylyshyn (1988) "Connectionism and Cognitive Architecture: A Critical Analysis" *Cognition* **28:** 3-71.

Garnham, Alan (1987) *Artificial Intelligence: An Introduction* (NY: Routledge).

Haugeland, John (1985) *Artificial Intelligence: The Very Idea* (Cambridge MA: MIT Press).

Lucas, J.R. (1961) "Minds, Machines, and Gödel" *Philosophy* **36:** 112-127.

McClintock, Alexander (1995) *The Convergence of Machine and Human Nature* (Brookfield VT: Avebury).

Morelli, Ralph, Miller Brown, Dina Anselmi, Karl Haberlandt, & Dan Lloyd (1992) *Minds, Brains, and Computers* (Norwood NJ: Ablex).

Penrose, Roger (1989) *The Emperor's New Mind* (NY: Oxford UP).

Penrose, Roger (1994) *Shadows of the Mind* (NY: Oxford UP).

Rapaport, William (1986) "Philosophy of Artificial Intelligence: A Course Outline" *Teaching Philosophy* **9.2:** 103-120.

Robinson, William (1992) *Computers, Minds and Robots* (Philadelphia: Temple UP).

Weizenbaum, Joseph (1976) *Computer Power and Human Reason* (San Francisco: Freeman Press).