THE QUESTION OF ARTIFICIAL INTELLIGENCE,
Brian Bloomfield, ed., Croom Helm, London, New York, Sydney.

This book belongs to a genre that treats a scientific field using various social science and humanistic disciplines, e.g. philosophy, history, sociology, psychology and politics. Scientists often complain about the results, both generally (judging the whole effort as wasted) and specifically (citing instances of ignorance and misunderstanding). I'm open minded about the general activity; maybe the sociology of research in AI has independent intellectual interest, though surely less than that of AI itself, and maybe sociological observations might cause participants in the field to change the way they do something, e.g. recognize achievement, define authority and distribute rewards. This review mainly concerns specific matters, and is mainly negative, complaining about ignorance and prejudice. The review also contains some suggestions about how this kind of thing can be done better — assuming it is to be done at all.

The successive chapters are entitled "AI at the Crossroads" by S. G. Shanker dealing with philosophy, "The Culture of AI" by B. P. Bloomfield, "Development and Establishment in AI" by J. Fleck, "Frames of AI" by J. Schopman, "Involvement, Detachment and Programming: The Belief in PROLOG" by P. Leith and "Expert Systems, AI and the Behavioural Coordinates of Skill" by H. M. Collins.

Reading "AI at the Crossroads" suggests entitling this review "Some Philosophers at a Crossroads". Shanker's path from the crossroads would lead to epistemology and the philosophy of mind leaving philosophy entirely. AI programs require knowledge and belief and their construction requires their formalization and scientific study. Shanker ignores this area in which philosophers and AI researchers have begun to co-operate and compete. Instead he considers the idea of artificial intelligence to be a "category error" of some almost unintelligible sort.

To someone engaged in AI research, it seems odd that for all his denunciation of AI, it isn't clear whether Shanker argues that there is any particular activity in which the external performance of computer programs must remain inferior to that of humans. It seems likely that he isn't making such a claim. Instead, much of what he says seems to be just an extreme demand that different levels of organization not be related in the same explanation. The most striking example of this is "... the psychologist can have no recourse to neural nets in order to explain, for example, the results of 'reaction time studies' ".

Shanker's 124 notes include no reference to the last 30 years of technical literature of AI, e.g. no textbook, no articles in *Artificial Intelligence* and no papers in the proceedings of the International Joint Conferences on AI. This permits him to invent the subject.

Thus he invents and criticizes an ideology of AI in which what a computer program knows is identified with the measure of information introduced by Claude Shannon in 1948. I wasn't aware that I or any significant AI pioneer made that identification, and it finally occurred to me to check whether even Shannon did. He didn't. His 1950 paper "Programming a Computer for Playing Chess" cited in Shanker's article never mentions information in the technical sense he introduced two years earlier.

While AI can only bandy words with Shanker and people in similar activity, we have serious business with many other philosophers. An intelligent program must have a general view of the world into which facts about particular situations fit. It must have views about how knowledge is obtained and verified. It must be able to represent facts about the effects of actions. It must have some idea of what choices are available to itself and other intelligences. This overlap in subject matter between AI and philosophy has led to increasing interaction.

Examples of philosophical work relevant to AI (besides mathematical logic) include the work of Frege (sense and denotation), Gödel (modern mathematical Platonism), Tarski (theory of truth), Quine (ontology and bound variables), Putnam (natural kinds), Hintikka (formalization of facts about knowledge), Montague (paradoxes of intensionality), Kripke (semantics of modality), Gettier (examples on intensionality), Grice (conversational implicatures), and Searle (performatives). However, all these topics need to be treated more modestly (in scope) and more formally and precisely than is usually done in philosophy. In addition to the aid AI has received from the above, we should also mention the encouragement received from Daniel Dennett.

In exchange, I believe that AI's concrete approach to epistemology will greatly affect philosophy. Indeed philosophers, e.g. Hintikka, and mathematical logicians are already studying the formalization of nonmonotonic reasoning, a topic originated in AI.

"The Culture of AI" argues that the ideas put forth by AI researchers (and scientists generally) should not be discussed independently of the culture that developed them. I don't agree with this, but have no objection to also discussing the culture. A rather extreme example of considering culture is favorably cited by Bloomfield, namely Athanasiou's

"The culture of AI is imperialist and seeks to expand the kingdom of the machine .... The AI community is well organized and well funded, and its culture fits its dreams: it has high priests, its greedy businessmen, its canny politicians. The U.S. Department of Defense is behind it all the way. And like the communists of old, AI scientists believe in their revolution; the old myths of tragic hubris don't trouble them at all".

It's rather hard to get down to discussing declarative vs. procedural representations or combinatorial explosion after such bombast. Moreover, whether current expert system technology is capable of writing useful programmed assistants for American Express authorizers, general medical practitioners, "barefoot doctors" in China, district attorneys or Navy captains is an objective question, and it doesn't seem that Bloomfield intends to help answer it.

We can't tell whether there is much to say about how the AI cultural milieu influenced its ideas, because Bloomfield's information about the AI culture is third hand. There is no sign that he talked to AI students or researchers himself. Instead he cites the books by Joseph Weizenbaum and Sherry Turkle. Weizenbaum dislikes the M.I.T. hackers, AI and otherwise; they don't like him either. He also confuses hackers with researchers; these groups only partly overlap. Turkle at least did some well prepared interviewing of both hackers and researchers. However, she doesn't make much of a case that the ideas stemmed from the culture *per se*. Indeed the originators of many of the ideas were and aren't participants in the informal culture of the AI laboratories. It occurs to me that since most of what we know about Socrates's ideas comes from Plato, perhaps the authors of this volume consider it unfair to use primary sources even in studying the activities of people alive and active today.

"Development and Establishment in AI" contains a lot of administrative history of AI research institutions and their government support. The information about Britain is moderately voluminous and seems more or less accurate, and the paper contains almost all the references to actual AI literature that occur in the volume.

Its American history is less accurate. There was no "Automata Studies" conference held in 1952. The volume of that title was composed of papers solicited by mail. The Dartmouth Summer Project on Artificial Intelligence was not a "summer school", i.e. the participants were not divided, even informally, into lecturers and students. The Newell-Simon group began its activities about two years before the Dartmouth conference. It is indeed true that the pioneers of AI in the U.S. met each other early, formed research groups that made continued contributions, and became authorities in the field. It's hard to see how it could have been otherwise. A fuller picture would also mention also-rans in the history of AI, people whose ideas did not meet with success or acceptance and dropped out.

The "AI establishment" owes little to the general "scientific establishment". AI would have developed much more slowly in the U.S. if we had had to persuade the general run of physicists, mathematicians, biologists, psychologists or electrical engineers on advisory committees to allow substantial NSF money to be allocated to AI research. Moreover, the approaches to intelligence originated by Minsky, Newell, Simon and myself were quite different from those advocated by Norbert Wiener, John von Neumann or Warren McCulloch.

Our good fortune with ARPA is due to its creation with new money at a time when we were ready to ask for support and very substantially to the psychologist J. C. R. Licklider. Licklider was on the Air Force Scientific Advisory Board around 1960 and argued that large command and control systems were being built with no support for the relevant basic science. ARPA responded by offering to create an office and budget for such support if Licklider would agree to head it. AI was one of the computer science areas Licklider and his successors at DARPA consider relevant to Defense Department problems. The scientific establishment was only minimally, if at all, consulted. In contrast European AI research long depended on crumbs left by the more established sciences. Recent PhDs were unable to initiate the research, and the European heads of AI laboratories often have been older people with existing reputations in other fields.

We make a final remark about the Lighthill report which initiated one of the dry periods in British AI funding. When a physicist is forced to think about AI he generally reinvents the subject in his individual way. Some expect it to be easy and others impossible. Lighthill was in the latter category. In the 1974 BBC debate, I thought I had a powerful argument and asked Lighthill why, if the physicists hadn't mastered turbulence in 100 years, they should expect AI researchers to give up just because they hadn't mastered AI in 20. Lighthill's reply, which BBC unfortunately didn't include in the broadcast, was that the physicists should give up on turbulence. Hardly any physicists would agree with Lighthill's statement, and maybe he didn't mean it.

Despite the deficiencies indicated above, the paper shows that attention to detail does pay off in useful information about history.

"Frames of Artificial Intelligence" by J. Schopman purports "to sketch a close-up of a crucial moment in the history of Artificial Intelligence (AI), the moment of its genesis in 1956". Schopman begins by telling us that "an exposition will be given of the investigative method used, SCOST — the 'Social construction of science and technology'." The "crucial moment" is stated to be the Dartmouth Summer Research Project on Artificial Intelligence except that Schopman refers to it as a conference and also mixes it up with the *Automata Studies* collection of papers. The papers for that collection were solicited starting in 1952, and the volume was finally published in 1956. The Dartmouth project did not result in a publication.

Whatever the SCOST method includes, it evidently doesn't include either interviewing the participants in the activity (almost all of whom are still alive and active) or looking for contemporary documents. The contrast with Herbert Stoyan's work on the history of the LISP programming language is amazing. Stoyan started his work while still living in Eastern Germany and unable to travel. Nevertheless, he wrote to everyone involved in early LISP work, collected all the documents anyone would copy for him and was able to confront what people told him in letters and interviews (after he was allowed to emigrate) with what the early documents said. He eventually came to know more about LISP's early history than any individual participant. If Schopman or anyone else wants to know what we had in mind when we proposed the Dartmouth study, he should obtain a copy of the proposal. If he wants to know why the Rockefeller Foundation gave us the $7500, he could begin by asking them if anyone there wrote a memorandum at the time justifying the support.

Old proposals and old granting-agency memoranda documenting their support decisions are an important unused tool in the recent history of science. The proposals often say in ways unrecorded in published papers what the researcher was hoping to accomplish, and the support memoranda tell what the agency thought it was accomplishing. Old referees' reports on papers submitted for publication and proposal evaluations provide another useful source. Were there referees' reports on Einstein's 1905 papers? In the U.S.A., the Freedom of Information Act provides an important way of find out what people in Government thought they were doing.

Now let's return to Schopman's actual speculations about what people were doing. He says that the Dartmouth "conference" was "a result of the choices made by a group of people who were dissatisfied with the then-prevailing scientific way of studying human behaviour. They considered their approach as radically different, a revolution — the so-called 'cognitive revolution'." Schopman has made all that up — or copied it from journalists who made it up.

The proposal for the Dartmouth conference, as I remember having written it, contains no criticism of anybody's way of studying human behavior, because I didn't consider it relevant. As suggested by the term "artificial intelligence" we weren't considering human behavior except as a clue to possible effective ways of doing tasks. The only participants who studied human behavior were Newell and Simon. Also, as far as I remember, the phrase 'cognitive revolution' came into use at least ten years later.

For this reason, whatever revolution there may have been around the time of the Dartmouth Project was to get away from studying human behavior and to consider the computer as a tool for solving certain classes of problems. Thus AI was created as a branch of computer science and not as a branch of psychology. Newell, Simon and many of their students work both in AI as computer science and AI as psychology.

Schopman mentions many influences of earlier work on AI pioneers. I can report that many of them didn't influence me except negatively, but in order to settle the matter of influences it would be necessary to actually ask (say) Minsky and Newell and Simon. As for myself, one of the reasons for inventing the term "artificial intelligence" was to escape association with "cybernetics". Its concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert (not Robert) Wiener as a guru or having to argue with him. (By the way I assume that the "Walter Gibbs" Schopman refers to as having influenced Wiener is most likely the turn-of-the-century American physicist Josiah Willard Gibbs, though possibly McCulloch's colleague Walter Pitts). Minsky tells me that neither Wiener nor von Neumann, with whom he had personal contact, influenced him, because he didn't agree with their ideas. He does mention influence from Rashevsky, McCulloch and Pitts.

Schopman paints a picture of the intellectual situation in 1956 based on the publications of many people who wrote before that year. Maybe that was the intellectual situation for many, but I suspect the situation was more fragmented than that; many people hadn't read the papers Schopman identifies as influential. For example, the idea that programming computers rather than building machines was the key to AI received its first public emphasis at the Dartmouth meeting. None of von Neumann (surprisingly), Wiener, McCulloch, Ashby and MacKay thought in those terms. However, by the time of Dartmouth, Newell and Simon, Samuel and Bernstein had already written programs. McCarthy and Minsky expressed their 1956 ideas as proposals for programs, although their earlier work had not assumed programmable computers.

However, Alan Turing had already made the point that AI was a matter of programming computers in his 1950 article "Computing Machinery and Intelligence" in the British philosophy journal *Mind.* When I asked (maybe in 1979) in a historical panel who had read Turing's paper early in his AI work, I got negative answers. The paper only became well known after James R. Newman reprinted it in his 1956 *The World of Mathematics.* Actual influences depend on what is actually read. A diligent historian of science could check what papers were referred to.

Finally, there is Schopman's chart that associates AI frames (paradigms) with periods. In no way did these "paradigms" dominate work in the periods considered. There have been, however, substantial shifts in emphasis at various times since the Dartmouth conference. Someone studying this will need to subdivide the AI "paradigm" in order to say which "subparadigms" were popular at different times. One way to study this would be to classify PhD theses and IJCAI papers and count them.

"Involvement, Detachment and Programming: The Belief in Prolog" by Philip Leith treats the enthusiasm for Prolog as a sociological phenomenon analogous to the 16th century Ramist movement in the logic and rhetoric of law. The Britannica article on rhetoric says the Ramist movement emphasized figures of speech. I wasn't convinced that this has much analogy to Prolog. Leith's complaint that Kowalski's work on expressing the British Nationality Act in logic programming was supported by the wrong Research Council leads this American to speculate that purely British quarrels about money and turf are being reflected; Americans should discreetly tiptoe from the room. At the 1987 Boston conference on AI and law, the Kowalski work was referred to respectfully by both the computer scientists and the lawyers present.

"Expert Systems, Artificial Intelligence and the Behavioural Co-ordinates of Skill" by H. M. Collins, a sociologist, is the paper admitting the most straightforward response. Collins classifies expert systems into four levels beginning with computerization of a rule book, followed by the incorporation of heuristics obtained by interviewing experts but used by humans only as an adviser, followed by expert systems acting autonomously and finally by systems with common sense. This seems like a useful classification along one dimension.

He also has nice examples. One concerns a referee's decision when one side in cricket inadvertently had an extra man on the field during an "over", and the fact wasn't noticed till much later. In deciding what to do the referee had to go beyond the rule book. Presumably he took at least the following considerations into account: his intuitive concept of fairness, the probable perceptions of fairness by the players, the spectators and his fellow officials, the need to keep the game going, maintaining the authority of the officiating system and the need to reach a prompt decision. All these considerations involve the referee's common sense and refereeing experience. None of them are in the rules of cricket, although some may be in books about refereeing or in a handbook for cricket referees. An AI system with human refereeing capability would need general common sense knowledge and reasoning ability. Collins's intuition and that of the other authors in this collection is that this is not possible.

AI has to take such examples as challenges. Should we be stumped, we should admit it for the time being and promise to tackle the problem later. However, I don't feel stumped by the cricket referee problem. I agree with Collins that the solution doesn't lie in simple extensions to the cricket rule book. This would indeed require an impractical or even impossible number of rules. However, the formalization of common sense is leading to ideas like formalized context with nonmonotonic rules about how contexts might be extended. These are discussed in (McCarthy 1979, 1986, 1987). These approaches are just beginning and took a long time to reach the concreteness required even to write papers. They still may not work.

However, it is not justified for philosophers or sociologists to claim to have shown that common sense can't be formalized. (The pioneer sinner in this respect was Wittgenstein). If you want to show something is impossible you have to prove theorems, as did Boltzmann (with thermodynamics), Gödel and Turing. Then you must be careful not to go beyond what the theorems say in your intuitive exposition.

Philosophers, etc. are entitled to their negative intuitions, but they should try to concretize them. For example, let them try to devise the easiest task that they think computers can't do. If they are willing to read current papers, they can be even more useful. They can try to devise the easiest problem the current AI methods can't do.

REFERENCES

McCarthy, John (1979): "First Order Theories of Individual Concepts and Propositions", in Michie, Donald (ed.) *Machine Intelligence 9*, (University of Edinburgh Press, Edinburgh).

McCarthy, John (1986): "Applications of Circumscription to Formalizing Common Sense Knowledge" *Artificial Intelligence*, April 1986

McCarthy, John (1987): "Generality in Artificial Intelligence", *Communications of the ACM.* Vol. 30, No. 12, pp. 1030-1035

McCulloch, W and Pitts, W. (1943): "A logical calculus of the ideas immanent in nervous activity".*Bulletin of Mathematical Biophysics*, 5, 115-137.

Shannon, C.(1950): "Programming a computer for playing chess". *Philosophical Magazine*, 41.

Turing, A.M. (1950): "Computing machinery and intelligence". *Mind*, 59, 433-60.

Wiener, N. (1948): Cybernetics. New York, Wiley.

John McCarthy
Computer Science Department
Stanford University
Stanford, California 94305