Discussion

The future of AI and human flexibility *

Manfred KOCHEN
Mental Health Research Institute, The University of Michigan, Ann Arbor, MI 48109, USA

Manfred Kochen was educated at MIT with a B.S. in Science (1950) and at Columbia University in New York, where he received an MA in 1951 and a Ph.D. in 1955 in Applied Mathematics.

Since 1972 he has been Professor of Information Science, with a joint appointment as Adjunct Professor of Computers and Information Systems at the Graduate School of Business Administration and as Research Scientist at the Mental Health Research Institute. His major research interests are in problems of representation of knowledge that could help people cope with a greater variety of tasks in more effective ways as well as in organization of knowledge. He is the author or editor of six books, including *Information for Action*, he has over 150 scientific papers in technical journals or books, including three on the process of referee-selection and peer review.

In 1974 he was awarded the Award of Merit by the American Society for Information Science. He was also named National Lecturer by the Association for Computing Machinery. He has been an editor of the *Journal of the ACM*, of *Behavioral Science* and is on the editorial boards of several international journals. He has been very active in the American Society for Information Science.

He is increasingly interested in all aspects of 'Coping with Social Complexity', which is the title of the section of *Human Systems Management* that he is responsible for as managing editor. He also has a long-term interest in the process of doing science, of which journal publication comprises the later stages. He is therefore receptive to and interested in good new ideas for innovations in the publication process leading toward improved papers and particularly toward helping authors to improve the quality of manuscripts they produce in the first place.

A sample of alternate futures of AI is used to illustrate a fruitful approach to meta-AI. At least two important issues generated by progress in AI require analysis. One is the trade-off between building an AI dedicated to a specialized, well-defined domain of problems, at which it excels, and building a flexible AI that can reprogram itself quickly to cope with a variety of unanticipated new tasks. Another central issue concerns the role and value of being human in the face of pressing change of the kind believed to be engendered by AI.

Keywords: Value of being human, flexibility, forecasting, prediction, robotics, conflict management.

* Presented in part before Michigan Society of Fellows, University of Michigan, Ann Arbor, March 15, 1984. Partial support under grant NSF IST-8301505 is gratefully acknowledged, as is the assistance of Charles Barr.

North-Holland
Human Systems Management 5 (1985) 159–162

1. Introduction

In a sense, the future of AI was in the late fifties. Herbert Gelernter's team at the IBM Research Center had constructed programs that could prove theorems in plane geometry [6], as expected of high school students, using heuristics based on the use of diagrams; some of the proofs it constructed surprised its originators to a greater extent than did the discoveries made by Lenat's AM [3]. Arthur Samuel developed a program that demonstrated learning from experience, using the game of checkers as a vehicle, by adjusting parameters in a polynomial that evaluated board positions and games in terms of piece count, control of the center, etc. [13]. He and I applied these techniques to the design of business games for use in training of executives [9]. Nils Aall Baricelli, in the early fifties at the Institute for Advanced Studies, simulated evolution using known genetic mechanisms such as cross-over, applied to 'numerical organisms' that resembled self-organizing geometric patterns [1]. Ray Solomonoff pioneered in the study of induction that led to what is now algorithmic complexity theory [14]. My own work on the formation of models and maps was the first to stress the representation issue. Marvin Minsky's 1961 bibliography on AI listed over 550 publications and my 1960 survey of cognitive mechanisms [8] listed about 200.

At least four approaches are used in thinking about the future. The first tries to *forecast* the most plausible course of events so that we can prepare. Solomonoff's commentary is an example. Though his expectations are enlightened and justified, other equally informed commentators may select a different sequence of milestones. An equally plausible forecast, for example, may be that Solomonoff's milestone E is a machine that has general problem-solving capacity near that of a soldier or officer in areas for which it has been designed – presumably in various tactical and strategic combat situations – rather than in mathematics, science and industrial science applications. Certainly there is great potential, interest and funding available for military applications of AI.

The second approach to futuristic thinking is to *predict*, with a view toward testing a theory, in this case a theory about the development of AI. Though some work on such models for the dynamics of scientific specialties has been started, it has not yet advanced to where scientifically testable predictions can be made. The use of citation analysis by Eugene Garfield [5] has identified a few research fronts based on co-citation clusters, such as (1) retrieval processes involving computational linguistics, (2) non-recursive grammars, (3) chemical structure elucidation, (4) knowledge engineering for medical diagnosis, and (5) theorem-proving. But this is retrospective rather than predictive.

The third approach to the future is to *shape* it according to a vision and an intention. This is the approach advocated by Minsky. It takes political and economic leverage as well as special entrepreneurial and managerial skills to do this. This approach is one of action rather than writing about it.

The fourth approach is to *imagine alternative futures* for scholarly analysis and to organize research and teaching. The purpose of this note is to illustrate that approach.

2. A way of thinking about AI

The ability to learn by acquiring and attaining knowledge and using reason to respond quickly and successfully to *new* situations is analogous to the ability to utilize energy for useful physical work. The latter involves conversion of low-grade forms of energy, such as heat, to higher-grade forms, such as mechanical or electrical energy. Possibly there is an analogous low-grade and high-grade knowledge. Von Bertalanffy [15] and Maslow have drawn a clear distinction between 'deficiency cognition', that is, cognition in the service of adapting and coping with reality, using perception and accepted symbolic frameworks of thought and language; and 'being cognition', peak experience ineffable in the sense that it cannot be transmitted using discursive forms. The former might be identified with the particulars of a concrete and ordinary experience, the latter with more abstract principles of some generality.

Corresponding to the different forms of energy we may have knowledge in the form of: (a) signals from an uncontrolled environment in which a robot is to act and by so doing, acquire coded representations of experiences or mathemas (that which is learned); (b) signals from a controlled environment in which a robot is to learn from a planned sequence of exposures; (c) know-how, ideas or heuristics in the minds of experts, tapped by knowledge engineers; (d) the materials (text and graphics) in encyclopedias; (e) the content of texts on specialized topics; (f) the content of articles at the front of an advancing research specialty; (g) the discoveries in the minds and notes of researchers prior to their expression in manuscript, tapped by science writers and close research peers. A more complete version of this list lends itself to morphological analysis, in which we ask how each form could be converted to all the others.

AI, interpreted as closer to knowledge *engineering* than to cognitive science, is a process of investigating how to solve problems. It is not a technology, which according to Anna Harrison is the process of producing and delivering goods and services. Nor is it a science, which she defines as the process of investigating physical, biological, behavioral, social economic, and political phenomena.

As a branch of engineering, the growth of AI is shaped by at least two forces: the pull of the market as it reflects societal needs, and creates niches; and the push of technology and technological innovation.

3. A small sample of alternative futures

3.1. Expert systems

Pull. Need for the services of many more experts, for example in medical specialties, than are available or being trained.

Push. Possibility of codifying existing expertise by tapping minds of experts, textbooks, selected articles with the help of software tools such as EMYCIN and computer facilities such as the SUMEX-AIM network.

Evaluation. There are not enough knowledge engineers capable of capturing these expertises, nor enough support tools, nor are enough being trained and developed. Hence researchers such as Lenat are attempting to automate both expertise and its conversion into an expert system, a difficult task.

3.2. Learning from world knowledge

Pull. Need for flexible robots such as those equipped with computer vision, capable of complex assemblies and operations of more than one kind, e.g. to dismantle unexploded bombs, flight fires, install many shapes and sizes of automobile bolts and screws.

Push. Advances in robotics fused with AI.

Evaluation. At a given level of expenditure, a robot may excel in a specified, specialized task such as dismantling a bomb, or it may be able to rapidly reprogram itself so as to switch from a bomb-dismantling task to fire-fighting, but it cannot sustain the knowledge base simultaneously enabling it to do both equally well.

3.3. Learning from sources of new knowledge

Pull. Need to cope with problems that existing knowledge may be relevant to *and* screening, evaluating, synthesizing the immensity of existing (and growing) knowledge so that it can be brought to bear where it is relevant and useful.

Push. Advances in information science beyond access to specified items.

Evaluation. This alternative may, like that of Minsky, be one that should be deliberately pursued with high priority.

3.4. Education under control of the learner

Pull. Need to obtain perspective, orientation, attitudes for acquisition of knowledge, know-how, skills, understanding, and wisdom for coping.

Push. The technology implicit in LOGO as advocated by Papert [12].

Evaluation. Also suitable for intentional pursuit; plagued by issue of who controls development.

4. Human flexibility

As suggested in alternative 3.2., there may be a basic principle to the effect that the depth of an intelligent system comes at the expense of its breadth and flexibility.

It is probably true that, to the extent that we humans (or a knowledge engineer observing us) can specify with sufficient precision what we do and how, we can program a computer to do it faster, more reliably, more cheaply. But we humans can do much more than we, or a knowledge engineer observing us, can describe.

The 'partnership' of robotics and AI advocated by Michael Brady [2] of the MIT Artificial Intelligence Lab would, he believes, bring manufacturing managers far closer to the dream of a versatile reprogrammable industrial robot that is a match for the reality of a high percentage of short runs in batch assembly processes. But a basic trade may have to be made. The price for not specifying in advance the special purpose of the system is that of building enough flexibility and adaptability into a general-purpose system. That price involves time and cost. In principle, it may be possible to build such a general purpose machine capable of rapid learning and self-adjustment. We humans seem to develop that way. However, the additional cost of building flexible machines to handle varied tasks, rather than using humans in such roles, may be great. In fact, it may be greater than the savings that can be achieved by using robots rather than humans for fixed, well-defined tasks.

Here, then, is a basis for a future division of labor between AI and humans. Let each person cope adequately with a variety of ill-defined tasks that cannot be pre-specified or anticipated and let AI perform prespecified tasks in prespecified domains of discourse. The latter may include much of what is now produced, and which provides people with gainful employment. But there is a vastly increasing plethora of societal problems – many of them generated by current technological advances – that involve tasks we cannot clearly anticipate, and seem now to require human attention, judgment, experience, and values for their resolution. The uniquely human reaction to such problems – at least in all our political and social history to date – is to generate 'solutions' which also become identified as problems at a later time. Far from being a vicious circle, this process exemplifies community learning and a form of 'progress' which we can realistically hope to manage and plan.

Examples: the transportation problem; the leisure pursuit problem; the genetic engineering problem; the poverty problem; the climate-change problem; the cultural heritage preservation problem; the aging problem; the land-use problem; the extraterrestrial contact problem; the participative political decision-making problem; the education problem; the problem of war; the genetic and ecosystem diversity problem; and the information overload problem.

Unfortunately, many aspects of working on these problems do not add economic value of the kind counted in other productive activities, and hence do not provide opportunities for gainful employment in present labor markets.

In the extreme case, where all labor needed in present technologies could be performed by non-human means, the wealth thus generated *must* be applied toward coping with these increasingly important and numerous problems for which there is now no labor market. It is essential, as noted by Nilsson in the article in this issue [11], that when we consider the dynamics of reallocating the wealth generated by automation that we ensure the appropriate direction and management of the transition.

5. Conclusion

I share with Joseph Weizenbaum [16] the concern that AI, to the extent that is possible to embody in a machine the ability to respond quickly and successfully to unanticipated new situations, may grow beyond human control and threaten human values. Perhaps the most optimistic outlook for the future is that the rapid development of applied AI at this time, which is likely to be followed by the search for more fundamental insights and radically new ideas about cognition, will spur fresh thinking about the value and role of humans. The value question may emerge as the central question of this century.

To objective forecasters, I suggest exploring the likelihood of military uses, leading toward the automated battlefield and the reactions this is likely to stimulate. Appropriate warnings should be sounded. To predictors, I suggest developing better models for the growth of AI. To shapers, I suggest developing fifth-generation languages, for international use in business, science and diplomacy, with capabilities for supporting conflict management along the lines of Fisher and Ury's *'Getting to Yes'* [4]. To analysts of alternate futures, I suggest studying the basic issues, such as the tradeoff between AI expertise in a specialized domain and a flexible, general-purpose AI. Above all, I suggest none of us lose sight of the ancient question of what it means to be free and human, of the value of being human, in the light of pressing changes and demands.

References

[1] Baricelli, Nils Aall, Symbiogenetic evolution processes realized by artificial methods, Revista Methods, 9 (1957) 35–36.

[2] Brady, Michael, Lecture presented at the University of Michigan, Artificial Intelligence Lab, Cambridge, MA, March 1984.

[3] Davis, Randall and Douglas Lenat, Knowledge-based systems in artificial intelligence (McGraw-Hill, New York, 1982).

[4] Fisher, Roger and William Ury, Getting to yes: Negotiating agreement without giving in (Houghton-Mifflin, Boston, MA, 1981).

[5] Garfield, Eugene, Citation analysis as a tool in journal evaluation, Science 178, 4060 (November 3, 1972) 471–479.

[6] Gelernter, Herbert and N. Rochester, Intelligent behavior in problem-solving machines, IBM J. of R&D 2, 4 (October 1958) 336–345.

[7] Harrison, Anna, Science, engineering, and technology, Science 223, 4636 (February 10, 1984) 534.

[8] Kochen, Manfred, Cognitive mechanisms, RAP-16 (IBM Research Center, Yorktown Heights, NY, 1960).

[9] Kochen, M. and A.L. Samuel, Business games for experimental research, IBM Research Planning Paper (1959).

[10] Minsky, Marvin, Toward a remotely-manned energy and production economy, Human Systems Management 5, 2 (1985).

[11] Nilsson, Nils, Artificial intelligence, employment, and income, Human Systems Management 5, 2 (1985).

[12] Papert, S., Mindstorms (Basic Books, New York, 1980).

[13] Samuel, Arthur, Some studies in machine learning using the game of checkers, IBM J. of R&D 3, 3 (July 1959) 221–229.

[14] Solomonoff, Ray, An inductive inference machine, IRE Convention Record (IT, August 1956).

[15] Von Bertalanffy, Ludwig, A systems view of man, in: Paul LaViolette, ed., A systems view of man: Collected essays (Westview Press, Boulder, CO, 1982).

[16] Weizenbaum, J., Computer power and human reason, (W.H. Freeman, New York, 1976).