Psychological Medicine, 1981, **11**, 449–453
Printed in Great Britain

EDITORIAL

Artificial intelligence[1]

Artificial intelligence (AI) is a fundamentally interdisciplinary subject, combining ideas from psychology, computer science, linguistics, mathematics and even philosophy (see, for example, Sloman, 1978). It is also a very young field. One consequence of its youth and interdisciplinary nature is that until fairly recently one of its most distinctive characteristics was its state of confusion. In the last few years, however, AI has begun to take on a much more coherent form. Reasons for this include the identification of key issues and fundamental areas and the development of approaches possessing considerable generality. I shall, very briefly, outline some of these below, but before doing so I shall explain why I believe AI to have a very important role to play in psychiatric research (indeed, also in psychological and sociological research and in the behavioural sciences generally).

The study of AI has two primary motivations. The first is to build computers (or, more correctly, to write programs) which behave in an intelligent way; and the second is to investigate the nature of human intelligence. Clearly, although these aims are related, they are not identical. In particular, we may note that while being an adequately close model of human brain function is a sufficient attribute for a program to demonstrate intelligent behaviour, it is not a necessary one. One can make an analogy with flying: while the study of flying animals (birds) can lead to ideas about flying (the use of wings), this is not a necessary approach (use rockets instead). Nevertheless, it is obvious that a mutual enrichment must follow from the interaction between AI and the human based mental sciences. Certainly, AI and cognitive psychology exhibit parallel theoretical developments and the hard formalism imposed by the rigorous theory formulation necessitated by the AI methodology has aided the testing of psychiatric (Colby, 1975) and sociological (Abelson, 1973) hypotheses. At present, both human-based models (for example, in natural language processing – see below) and abstract mathematical approaches (for example, in pattern recognition, Duda & Hart, 1973) have been successful. Some, however (for example, Hayes-Roth, 1978), believe that ultimately the best approach to AI may not be via the non-human information processing paradigms but via closer modelling of human memory and cognition. And, conversely, it may also be the case that the best approach to the study of human mentation is via the new field of AI. The point is that it is all very well formulating psychological and psychiatric theories verbally but, when using natural language (even technical jargon), it is difficult to recognize when a theory is complete; oversights are all too easily made, gaps too readily left. This is a point which is generally recognized to be true and it is for precisely this reason that the behavioural sciences attempt to follow the natural sciences in using 'classical' mathematics as a more rigorous descriptive language. However, it is an unfortunate fact that, with a few notable exceptions, there has been a marked lack of success in this application. It is my belief that a different approach – a different mathematics – is needed, and that AI provides just this approach (see also Boden, 1977). When formulating a theory as a program, which is exactly what AI does, the oversights are spotted, the gaps made apparent, and the oversimplifications made explicit. And a theory so formulated can be rigorously tested and appropriately refined. Weizenbaum (1976) puts it thus:

> The very eloquence that natural language permits sometimes illuminates our words and seems (falsely, to be sure) to illuminate our undeserving logic just as brightly. An interpreter of programming language texts, a computer, is immune to the seductive influence of mere eloquence.

It is, however, equally important to emphasize the qualitative distinction between the AI approach and the 'classical' mathematical or statistical approach. The former permits rigorous formulation

[1] Address for correspondence: Dr D. J. Hand, Biometrics Unit, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF.

and testing of hypotheses concerning the underlying psychological motivations (etc.), and not merely superficial ('mechanical', 'trivial') and easily observed variables.

In the remainder of this editorial I shall mention a few of the areas of research within AI which have been recognized as of fundamental importance, and also indicate a few of the many applications for which programs have been implemented. Thus, in view of the significant implications (social, philosophical, and so on), I shall not constrain myself merely to those formulations which are explicit models of human information processing. I have also attempted to refer mainly to key articles or review papers.

The single most important issue is probably that of knowledge representation: that is, the issue of how to describe or encode knowledge or information. (Here the parallel with certain areas of psychology is obvious.) A particular kind of behaviour may be almost impossible to model in one type of representation but trivial in another. Having said this, however, there are obvious advantages to be gained from general representations, representations which can be applied usefully in a large number of areas. One such is the concept of the 'frame'. This was first developed at the Massachusetts Institute of Technology (MIT) in the context of vision but has also been applied extremely effectively in the domain of natural language processing. A key paper is one by Minsky (1975), who states:

> A *frame* is a data-structure for representing a stereotyped situation like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.

As the focus of attention moves from one situation to another, so different frames are implemented. More correctly, different frames are *instantiated* (Bobrow & Brown, 1975; Kuipers, 1975): detailed information from the new situation is slotted into the chosen frame, replacing default values by values obtained from the currently experienced environment, be it verbal, visual, or in some abstract universe. Since the extra information may itself be in the form of frames, quite complex structures can be built up. The frame concept has led to the development of special high-level frame representation programming languages (see, for example, Bobrow & Winograd, 1977; Hayes, 1980). Related ideas include those of the script (Schank & Abelson, 1977; Lehnert, 1980) and the plan (Abelson, 1975; Schank & Abelson, 1975).

Given that a particular type of behaviour may appear simple or difficult to model, according to the representation chosen to realize it, the fact that it is not always obvious which of several competing representations is the best has led to some controversy. A good example of this is the fundamental choice between the opposing epistemologies of representing knowledge as procedures (i.e. program segments) or as declarations (i.e. as facts in a database): as Winograd (1975) puts it, choosing between 'knowing how' and 'knowing that'. The declarative approach, by separating the 'that' from the 'how' and concentrating on the former, leads to independence of distinct facts and hence to generalizability, modularity, and comprehensibility. Conversely, the procedural approach yields the capability for very rich interaction. While it is true that the two representations are trivially isomorphic at some level, the question is: what are the advantages to be gained by adopting one or the other (or a compromise) viewpoint? Frequently (always, some would say), either can be chosen for a particular domain and it is not clear which is the more efficient. Examples of the proceduralists' standpoint are given by Hewitt *et al.* (1973) and Winograd (1972), and of the declarativists' view by McCarthy & Hayes (1969). A thoughtful outline of the difference is provided by Winograd (1975) and an amusing 'imaginary, yet somehow representative conversation' from the procedural–declarativist feud appears at the end of Winston (1977).

Another related and important example of a fundamental choice in representation is that between conventional computer programs which operate 'on expected data in known formats using a prespecified and inflexible control structure' and *pattern-directed* systems in which parts of the program are activated whenever certain patterns or structures occur in the data (Waterman &

https://doi.org/10.1017/S0033291700052752 Published online by Cambridge University Press

Hayes-Roth, 1978*b*). Pattern-directed inference systems and *production systems* are being used more and more often in practical applications and I shall list some below.

Representation is one issue which has been identified as being of fundamental importance. Another is that of *search*. In the case of vision or natural language processing, for example, incoming information has to gain access to and interact with the appropriate part of the model, so that some efficient way has to be used to locate this part. Or consider the game of chess. In principle, one could list all possible games and choose the best move from a particular position by studying this list. In practice, this is inconceivable for even the fastest computer imaginable, so some simplified search method has to be found for identifying a good move in a reasonable amount of time. Once again, both abstract mathematics and the emulation of hypothesized human abilities have led to methods (Nilsson, 1971; Newell & Simon, 1972; Knuth & Moore, 1975). An interesting discovery of search theory, which seems at first counterintuitive, is the advantage to be gained by additional complexity. Waltz (1975), for example, shows how the introduction of the apparently additional complication of shadows into scenes makes it easier rather than harder to interpret the scene (the introduction imposes extra constraints on the number of possible interpretations).

It will be evident from the preceding brief sampling of some important issues that certain areas of study crop up again and again. Particular examples are the topics of natural language processing and vision. It is perhaps hardly surprising that natural language understanding has attracted a lot of interest since it can be argued that the same issues are involved in this limited domain as are involved in more general problems of intelligence, and it is also obvious that language understanding systems have direct practical application. Once again, many different approaches have been studied and a vast literature has accumulated. Examples of two different approaches are those of Winograd (1972) and Weizenbaum (1965). The former uses an internal representation which enables it to recognize the identity of paraphrases, to understand anaphora, and generally 'to infer the meaning of questions and commands that would otherwise remain opaque' (Boden, 1977) about its own (limited) universe. Weizenbaum's program uses pattern-matching processes and applies simple transformations to input sentences (so it can 'recognize' but not 'understand'). While the pattern-matching approach is not powerful enough to demonstrate the subtleties of parsing and comprehension exhibited by the knowledge-based approach, it does have the advantage that, whenever the input sentence contains a recognizable pattern, a response of some kind is given. Important recent developments include attempts to integrate the two approaches (for example, Schank, 1972, 1973; Wilks, 1972, 1976). It may be noted that in testing theories of language it is possible to adopt different viewpoints: Lehnert (1978) takes question-answering as the main aim, while Davey (1978) concentrates on discourse production. Examples of frame-based text processing are given by Cullingford (1977) and Rosenberg (1980). Further, different grammatical theories can be used (Winograd, 1972, for example, uses Halliday's (1970) 'systemic' grammar rather than Chomsky's syntactic theory).

Like language, vision is an area of immediate practical relevance (we have all heard how the latest generation of industrial robots have TV camera eyes to guide their work). Mathematical approaches (for example, Roberts, 1965) have been used, as also have more psychological based theories. Winston (1975) provides an excellent introduction. One thing which has become evident as understanding has grown is that much computation occurs in natural vision systems. This is only made feasible by massive computational parallelism – which is again being emulated in artificial systems (such as the CLIP4).

Natural language and vision are limited domains which have obvious relevance both to theoretical understanding and to practical applications. Another limited domain, but one which does not have any obvious practical relevance, is the development of games-playing programs. In fact, many of the fundamental ideas of AI have been developed in the game arena – and I can do no better than quote Clarke (1977):

> Many of us would maintain that chess, with its simple representation yet deep structure and richly developed culture, may even be the best system in which to study these problems (that is, AI) in their purest and most readily quantifiable form.

Recently, a new branch of AI has become important, namely, the *expert system* (Waterman & Hayes-Roth, 1978*a*; Michie, 1979, 1980). These are systems which hold representations of a particular limited domain and can respond intelligently and easily to questions and suggestions. They can thus act as consultants. I shall illustrate with a few medical examples, though such systems have been implemented in many different fields. MYCIN (Shortliffe, 1976) identifies bacteria in blood and urine samples and prescribes appropriate antibiotic treatments; INTERNIST (Pople *et al.* 1977) yields diagnoses in internal medicine (within their limited domains MYCIN and INTERNIST perform better than clinical consultants); PUFF (Kunz, 1978) diagnoses pulmonary function disorder; VM (Fagan, 1978) gives advice on the mechanical ventilation of patients in intensive care. An interesting spin-off of expert systems (perhaps one to be expected in the light of the comments above) is that the formalization of the domains as computer programs has led to improved understanding and clearer explanations of the subject matter. Michie (1980) calls this *knowledge refining*.

It is implicit in the above that such systems can improve their knowledge base by interacting with a human tutor. It is not so obvious that artificial systems can generate original material. Lenat (1976), however, has developed a system that searches for potentially interesting mathematical theorems (as distinct from other systems, which prove theorems). Boden (1977) also makes some telling points about creativity in AI. Research in other areas within AI has resulted in programs which can do geometric analogy tests as well as humans, perform symbolic integration better than humans, beat the world backgammon champion (Berliner, 1980), and make moves which even world champion chess players initially supposed to be poor until they had thought about it overnight.

We commented above that AI is a young discipline – and as with any young organism it is growing very rapidly. (There are already several journals devoted solely to AI research.) The reason for its youth is simply that effective experimentation in AI is entirely dependent on electronic computers, and these are a relatively recent development. It is even more recently that the importance of *user-friendliness* has been recognized. The expert systems discussed above do not give cursory diagnoses based on statistical analyses of databases: they arrive at their conclusions by interacting (usually in some subset of natural language) with the user. If asked why they have made an assertion they can answer – in as much detail as the user requires (see also Fox *et al.* 1980). One's thoughts inevitably turn to psychiatric interviews and diagnosis. To some, the very idea is immoral (Weizenbaum, 1976). But others (Colby, 1967) point out that if computers are helpful and if staffing problems mean there are too few psychiatrists, then we really have no choice. One must also remember the evidence that people prefer computer interviews when delicate or anxiety-causing topics are being discussed (Walton *et al.* 1973; Lucas *et al.* 1977).

These points are, however, quite distinct from that which I made at the beginning. Namely, that when we view AI as modelling man it follows a basically humanist approach, but one which nevertheless permits formal hypothesis formulation and testing.

D. J. HAND

REFERENCES

Abelson, R. P. (1973). The structure of belief systems. In *Computer Models of Thought and Language* (ed. R. C. Schank and K. M. Colby), pp. 287–340. W. H. Freeman: San Francisco.

Abelson, R. P. (1975). Concepts for representing mundane reality in plans. In *Representation and Understanding* (ed. D. G. Bobrow and A. Collins), pp. 273–309. Academic Press: New York.

Berliner, H. (1980). Computer backgammon. *Scientific American* **242**, 54–62.

Bobrow, D. G. & Winograd, T. (1977). An overview of KRL. *Cognitive Science* **1**, 3–46.

Bobrow, R. J. & Brown, J. S. (1975). Systematic understanding: synthesis, analysis, and contingent knowledge in specialised understanding systems. In *Representation and Understanding* (ed. D. G. Bobrow and A. Collins), pp. 103–130. Academic Press: New York.

Boden, M. (1977). *Artificial Intelligence and Natural Man.* Harvester Press: Brighton.

Clarke, M. R. B. (ed.) (1977). *Advances in Computer Class – I.* Edinburgh University Press. Edinburgh.

Colby, K. M. (1967). Computer simulation of change in personal belief systems. *Behavioural Science* **12**, 253.

Colby, K. M. (1975). *Artificial Paranoia.* Pergamon: New York.

Cullingford, R. C. (1977). Script application: computer understanding of newspaper stories. Research Report no. 116, Dept. of Computer Science. Yale University: New Haven, Ct.

Davey, A. (1978). *Discourse Production.* Edinburgh University Press: Edinburgh.

Duda, R. & Hart, P. (1973). *Pattern Classification and Scene Analysis.* Wiley: New York.

Fagan, L. M. (1978). Ventilator management. A program to provide on-line consultative advice in the intensive care unit. Memo no. 78-16. Stanford University Department of Computer Science: Stanford, Ca.

Fox, J., Barber, D. & Bardhan, K. D. (1980). Alternatives to Bayes? A quantitation comparison with rule-based diagnostic inference. *Methods of Information in Medicine* **19**, 210–215.

Halliday, M. A. K. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language* **6**, 322–361.

Hayes, P. J. (1980). The logic of frames. In *Frame Conceptions and Text Understanding* (ed. D. Metzing), pp. 46–61. De Gruyter: Berlin.

Hayes-Roth, F. (1978). Implications of human pattern processing for the design of artificial knowledge systems. In *Pattern Directed Inference Systems* (ed. D. A. Waterman and F. Hayes-Roth), pp. 333–346. Academic Press: New York.

Hewitt, C., Bishop, P. & Steiger, R. (1973). A universal modular ACTOR formalism for artificial intelligence. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pp. 235–245. Stanford, Ca.

Knuth, D. E. & Moore, R. W. (1975). An analysis of alpha–beta pruning. *Artificial Intelligence* **6**, 293–326.

Kuipers, B. J. (1975). A frame for frames: representing knowledge for recognition. In *Representation and Understanding* (ed. D. G. Bobrow and A. Collins), pp. 151–184. Academic Press: New York.

Kunz, J. (1978). A physiological rule-based system for interpreting pulmonary function test results. Memo HPP no. 78-19. Stanford University Department of Computer Science: Stanford, Ca.

Lehnert, W. G. (1978). *The Process of Question Answering.* Laurence Erlbaum: Hillsdale, N.J.

Lehnert, W. G. (1980). The role of scripts in understanding. In *Frame Conceptions and Text Understanding* (ed. D. Metzing), pp. 79–95. De Gruyter: Berlin.

Lenat, D. (1976). AM. an artificial intelligence approach to discovery in mathematics as heuristic search. Memo AIM no. 286. Stanford University AI Lab.: Stanford, Ca.

Lucas, R. W., Mullin, P. J., Luna, C. B. X. & McInroy, D. C. (1977). Psychiatrist and a computer as interrogators of patients with alcohol-related illnesses: a comparison. *British Journal of Psychiatry* **131**, 160–167.

McCarthy, J. & Hayes, P. (1969). Some philosophic problems from the standpoint of AI. In *Machine Intelligence* vol. 4 (ed. B. Meltzer and D. Michie), pp. 463–502. Edinburgh University Press: Edinburgh.

Michie, D. (1979). *Expert Systems in the Micro-electronic Age.* Edinburgh University Press: Edinburgh.

Michie, D. (1980). Expert systems. *The Computer Journal* **23** 369–376.

Minsky, M. (1975). A framework for representing knowledge. In *The Psychology of Computer Vision* (ed. P. H. Winston), pp. 211–277. McGraw-Hill: New York.

Newell, A. & Simon, H. A. (1972). *Human Problem Solving.* Prentice-Hall: Englewood Cliffs, N.J.

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

Pople, H. E., Myers, J. D. & Miller, R. A. (1977). DIALOG: a model of diagnostic logic for internal medicine. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Pittsburgh, USA.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In *Optical and Electro-optical Information Processing* (ed. J. T. Tippett, D. A. Berkowitz, L. C. Clapp, C. Clapp, C. J. Koester and A. Vanderburgh), pp. 159–198. MIT Press: Cambridge, Mass.

Rosenberg, St T. (1980). Frame-based text processing. In *Frame Conceptions and Text Understanding* (ed. D. Metzing), pp. 96–119. De Gruyter: Berlin.

Schank, R. C. (1972). Conceptual dependency: a theory of natural language understanding. *Cognitive Psychology* **3**, 552–631.

Schank, R. C. (1973). Identification of conceptualisation underlying natural language. In *Computer Models of Thought and Language* (ed. R. C. Schank and K. M. Colby), pp. 187–248. W. H. Freeman: San Francisco.

Schank, R. C. & Abelson, R. P. (1975). Scripts, plans and knowledge. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, pp. 151–157. Tbilisi, USSR.

Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding.* Laurence Erlbaum: Hillsdale, N.J.

Shortliffe, E. H. (1976). *Computer-based Medical Consultations: MYCIN.* Elsevier/North Holland: New York.

Sloman, A. (1978). *The Computer Revolution in Philosophy.* Harvester Press: Brighton.

Walton, H. J., Kiss, G. R. & Farvis, K. M. (1973). A computer program for the on-line exploration of attitude structures of psychiatric patients. MRC Speech and Communication Research Unit: Edinburgh.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. *The Psychology of Computer Vision* (ed. P. H. Winston), pp. 19–92. McGraw-Hill: New York.

Waterman, D. A. & Hayes-Roth, F. (eds.) (1978*a*). *Pattern Directed Inference Systems.* Academic Press: New York.

Waterman, D. A. & Hayes-Roth, F. (1978*b*). An overview of pattern-directed inference systems. In *Pattern Directed Interference Systems* (ed. D. A. Waterman and F. Hayes-Roth), pp. 3–22. Academic Press: New York.

Weizenbaum, J. (1965). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9**, 36–45.

Weizenbaum, J. (1976). *Computer Power and Human Reason.* W. H. Freeman: San Francisco.

Wilks, Y. A. (1972). *Grammar, Meaning, and the Machine Analysis of Language.* Routledge and Kegan Paul: London.

Wilks, Y. A. (1976). *Natural Language Understanding Systems Within the AI Paradigm.* Edinburgh University AI Dept: Edinburgh.

Winograd, T. (1972). *Understanding Natural Language.* Academic Press: New York.

Winograd, T. (1975). Frame representation and the declarative/procedural controversy. In *Representation and Understanding* (ed. D. G. Bobrow and A. Collins), pp. 185–210. Academic Press: New York.

Winston, P. H. (1975). *The Psychology of Computer Vision.* McGraw-Hill: New York.

Winston, P. H. (1977). *Artificial Intelligence.* Addison-Wesley: London.

https://doi.org/10.1017/S0033291700052752 Published online by Cambridge University Press