> The packaging for ignorance by artifice to obtain finance distorts not only the returns actually received by investors. Distortion of research activity in order to obtain finance on such terms is both a professional and a social cost that we in this country seem ill equipped to appreciate.

The main question to be asked of this book is, "Who is it for?". As the editors point out in the last chapter, the agreements and differences of opinion expressed at the colloquium have been preserved in the book. The transcripts of the informal discussion sessions at the end of each section are particularly effective in this connection. Great stuff for the practising researcher, who will be intimately familiar with the elusive issues involved, and who will thus be in a position to read between the lines. However, the discussions, together with the editors' own contributions, give the book a tone of superficiality and off-handedness which is dangerous and inappropriate for a more general audience. In spite of this, some of the individual chapters are particularly sound, and deserve wider exposure.

I want to see technical managers and engineers becoming acquainted with the issues covered in this book. But I would tell them not to believe everything they read in it. The simplistic, shoot-from-the-hip flavour may be instinctively appealing, but it endangers credibility, and should not be interpreted as solid worldly-wise experience. Despite Winston's "Six Ages of Artificial Intelligence", 20 years is not a very long time. What we need, above all is a little patience.

KNOWLEDGEABLE MACHINES

Knowledge-based Systems in Artificial Intelligence.
Randall Davis
Douglas Lenat.
McGraw-Hill: 1983. Pp. 490.

John Fox, Imperial Cancer Research Fund

Knowledge engineering is all the rage, just as genetic engineering was a few years ago. There are many similarities between these two new fields. Both have spun off from relatively recent sciences, molecular biology and artificial intelligence. Both attract the attention of the public and of investment managers to an almost embrassing degree. Both have yet to supply practical products on a large scale.

Fashion is fickle, however, and one might be forgiven for thinking that the knowledge engineering bubble will burst, or that the field will mature in the usual way and be rewarded with a respectable but properly modest place in the scientific establishment. Knowledge engineers, of course, do not believe that this will be the fate of their subject. Indeed, devotees of the parent discipline, artificial intelligence (AI), seem to have higher aspirations than merely "a proper place in the scientific establishment" (though the contributor to a television programme who said that "the three most important events in the history of the Universe were The Creation, the appearance of life, and the discovery of artificial intelligence" must have made even the most flamboyant AI scientist wince a little).

Given such high embition, I approached the volume by Davis and Lenat with the questions of scientific and engineering credibility in mind. It seemed promising. Both authors did the work described in the book as PhD dissertations at the Stanford University Heuristic Programming Project, still perhaps the leading centre of knowledge engineering. Furthermore the book consists of two complementary monographs; one with a scientific flavour and one with more of an engineering bias.

Douglas Lenat sees his work on AM (the Automated Mathematician) as scientific research with engineering implications. His efforts centred on the concept of "discovery", and the extent to which interesting discoveries in number theory can be made using programs that can use simple but powerful rules to generate and evaluate large sets of possible conjectures and hypotheses. For example "if $f(x, y)$ is interesting then $f(x, x)$ may also be interesting". If f is ADD, this suggests the concept of doubling, if TIMES then squaring, and so forth. Lenat gives many examples of more sophisticated "heuristics". The hypothesis that led to AM is that discovery (invention, creativity...) should not be assumed to be a mysterious mental faculty, but is the product of mental information processing mechanisms which are subject to logical analysis and thus to computer simulation. AM produced a few ideas which were new to Lenat, a couple that were "new to the world" and several "...additional 'new' concepts were created which both AM [and Lenat] agreed were better forgotten".

Lenat's work does not show that heuristic search is all there is to human discovery, only that there is a prima facie case that computers can be inventive. His hypothesis that discovery could be achieved by a physical device is not new — nothing is — but the demonstration that heuristics can lead to interesting discoveries, now, is dripping with provocative questions.

However, although Lenat clearly feels strongly about the importance of heuristic discovery processes, he is less than clear about the conceptual roots of the method. For example the "interestingness" of a conjecture is crucial to AM's performance, because it controls which of a prodigious set of conjectures will be explored further. Lenat is quite candid that the formula which the program uses to calculate "interestingness" is *ad hoc* — "Since it has worked successfully, I have not messed with it", though he adds subsequently that "...recent experiments tend to confirm that the particular form of the formula is unimportant". Nor does Lenat make it clear what he believes the philosophical status of his achievement to be. He says AM is interesting because it works, but he does not discuss whether heuristic discovery is the (or a) mechanism of human discovery or whether there are likely to be other interesting classes of discovery principle. Here he manifests an engineering reserve rather than a scientific curiosity. An important point I could not find addressed is whether AM is an artifact — successful only because it builds on generations of mathematicians who created the conceptual framework within which the heuristics can be expressed. This need not make Lenat's achievement less interesting but it would perhaps make it less significant.

Much of the current enthusiasm for knowledge engineering has been triggered by rapid growth of the belief that if it is not practical already it will be next week. The appearance of "expert systems" — systems which resemble human experts in carrying out specialized tasks — is largely responsible. One of the earliest and best understood was the MYCIN system which was developed to assist doctors in management of bacterial infections. By conventional standards MYCIN is a large program. Much of its size is due to the large body of knowledge it contains about infecting organisms, how they manifest themselves and how they are treated. It appears that the performance of expert systems is determined more by how much knowledge the system has than by the sophistication of the techniques that exploit the knowledge to recommend decisions.

The difficulty in compiling large, comprehensive bodies of knowledge is widely considered to be the main obstacle to commercial explotation of expert systems. Texbooks are unsuitable and just questioning human experts is not systematic enough. Consequently there is great interest in methods that can speed up the acquisition of knowledge bases. Randall Davis's "Teiresias" grew out of the MYCIN project; it is a program that explores the potential of partially automating the acquisition and refinement of knowledge by interacting with a human expert. Teiresias is named after the blind seer in Sophocles' "Oedipus the King" *(sic)*.

Davis distinguishes three types of knowledge. "Object knowledge" refers to the basic application concepts, e.g. medical knowledge of organisms, cultures, drugs etc. "Conceptual Knowledge" describes these objects in more general terms; drug dosage is a quantitative concept with values in a particular range, certain organisms have typical features, unique features and so forth. Finally there is "meta-knowledge" or knowledge about how to acquire object and conceptual knowledge from the human expert and represent it in forms that an expert system can reason with.

One aim of the Teiresias project was to refine this idea of meta-knowledge and build an expert system whose expertise is in the construction of other expert systems. Teiresias operates by repeatedly testing the developing expert system on sample problems given by the expert. When the system goes wrong it focuses attention on missing or incorrect information through an elaborate question-answer dialogue. Teiresias helps to focus on the substance of the knowledge that is needed rather than on how the expert system should be programmed, ensuring that the knowledge is fully specified and reminding the experts of all the items that must be supplied and in what form.

A particularly useful section of the book deals with how meta-knowlede can be used to provide explanations of the applications system's reasoning about a problem. This is important for a designer trying to understand why an unfinished system is misbehaving (due to missing or incorrect knowledge). However the facilities are also available in the finalized system and therefore can be used to understand the expert system's recommendations during practical use. This ability to give explanation accounts for much of the interest by software engineers in expert system techniques, and Davis gives a clear account of the techniques used in providing explanations, the concepts they are based upon and a straightforward summary of their limitations.

As a whole, the book is an impressive record of two novel pieces of work in artificial intelligence and knowledge engineering. Both monographs provide valuable information for other designers and research workers. Unfortunately, however, the book is not easy to read or easy to use. Minor subjects are dealt with in as great detail as important ones; the text is heavily punctuated with difficult notations and abbreviations; there is little cross-linkage between monographs; and there is no index. The authors are also somewhat inward looking, giving little consideration to previous thinking about knowledge or scientific discovery outside AI, or to the massive body of research on decision support which has not employed knowledge-based concepts. So while their book will be compulsive reading for the committed specialist, I fear that it will fail to convince many people who are trained in other philosophies and traditions.

Davis and Lenat have so much to say that it is a pity that they didn't say it more accessibly. To be fair their

work is more briefly written up elsewhere in technical journals, but these are for an informed readership. The book was an opportunity to provide a distillation of profound ideas for a broadly based audience of scientific peers. The authors chose to write a tome.

SEARCH FOR COGNITIVE SCIENCE

Conceptual Structures: Information Processing in Mind and Machine.
By John F. Sowa.
Addison-Wesley: 1984. Pp. 481.

John Fox, Imperial Cancer Research Fund

Artificial intelligence has produced the latest new technology to enter popular awareness. "Expert systems" are advanced computer systems which solve problems that were previously thought to be the preserve of human beings; not just any human being, but experts. Artificial intelligence (AI) techniques are used to emulate (to a first approximation) the problem-solving and decision-making of the expert human mind. Systems for medical diagnosis, geological data interpretation, semi-conductor chip design, fertilizer selection, tax advice and so on have all been successfully developed to some degree, and the potential range of this new technology is enormous — even the emulation of exotica such as the military, scientific or political mind no longer seems entirely absurd.

In the middle of all this enthusiasm for "knowledge engineering" it is easy to forget that artificial intelligence is not new, and that it did not originally set out to be an engineering discipline but a science. The early workers defined their goal to be an understanding of the principles that underly intelligent behaviour in man, animals and (their special contribution) machines.

Artificial intelligence emerged in its modern shape when computers became available during the mid-1960s to scientists of many disciplines. Intelligence was seen as the capacity to manipulate symbols and, consequently, it could be understood in terms of symbol-manipulating models run on computers. Since the very early work on problem-solving and theorem-proving, researchers in AI have gone on to model the processes of visual perception, the understanding of language and speech, learning and the acquisition of intellectual skills, and even the formation of beliefs, social attitudes and the development of psychiatric conditions.

These subjects somewhat resemble those which have traditionally been the province of psychology. So what were psychologists doing during this period? Partly due to its own successes, and partly due to the influence of AI concepts, "cognitive psychology" was also developing rapidly in such fields as perception, memory, decision-making and so on. The relationship between the two communities was usually fraternal. The reference sections of psychological papers often acknowledged the contributions of AI, and vice versa.

Often, however, the acknowledgements were superficial only, and discussions were peppered with critical exchanges. Programmers in AI may be vigorous and inventive, say many cognitive psychologists, but they are not *scientists* with the normal discipline for careful, systematic enquiry that the term implies. The counter is that psychologists are obsessed with methodology. They miss the point, and the insights, of AI demonstrations of what is possible with symbol-processing mechanisms.

Over the past ten years there have been a few, but generally less than successful attempts at synthesis, although a name has been invented, "cognitive science", which is supposed to cover both scientific AI and cognitive psychology. Unfortunately most cognitive scientists are only well informed about one of the branches. They are also trained very differently and their books and journals show it. It is against this background that John Sowa's book appeared.

At first sight the book looks as though it might be something special indeed. The chapter headings cover a range of core ideas, including the philosophical basis of AI, reasoning and computation, language, psychological evidence and — above all — knowledge. Furthermore the preface declares that the book "combines AI and cognitive science approaches... In combining insights from each of the separate fields, the book gives a unified view of knowledge representation".

Sowa applies a general plan to each chapter wherever possible, the first few pages being background material containing wide-ranging discussion. These sections usually summarize psychological research and findings (in perception, language and so on) before going into more detailed, technical presentations on the same subject where the emphasis is on computation and formalization. Sowa maintains an attractive, readable style throughout the book, studded with illuminating illustrations of his points from a remarkable variety of fields ranging from optical illusions perceived by the peasants of Uzbekistan to the structure of early Greek poetry.