Computers and Thought

A Practical Introduction to Artificial Intelligence

Mike Sharples,
David Hogg,
Chris Hutchinson,
Steve Torrance, and
David Young

Computers and Thought

A Practical Introduction to Artificial Intelligence

Mike Sharples,
David Hogg,
Chris Hutchinson,
Steve Torrance, and
David Young

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Mike Sharples ... [et al.].
Computers and thought: a practical introduction to artificial intelligence

p. cm. — (Explorations in cognitive science 5)
"A Bradford book."
Bibliography: p.
Includes index.
1. Artificial Intelligence. 2. Cognitive Science. I. Sharples, Mike, 1952 – . II. Series.
Q335.C57 1989   006.3 — dc19   89-2541
ISBN 0-262-19285-3   CIP

Explorations in Cognitive Science
Margaret A. Boden, Editor

1. *Mental Processes: Studies in Cognitive Science,*
   H. Christopher Longuet-Higgins

2. *Psychosemantics: The Problem of Meaning,*
   Jerry A. Fodor

3. *Consciousness and the Computational Mind,*
   Ray Jackendoff

4. *Artificial Intelligence in Psychology: Interdisciplinary Essays,*
   Margaret A. Boden

5. *Computers and Thought: A Practical Introduction to Artificial Intelligence,*
   Mike Sharples, David Hogg, Chris Hutchinson, Steve Torrance, and David Young

The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

Alan Turing, in
"Computing Machinery and Intelligence"

Preface

The aim of this book is to introduce people with little or no computing background to artificial intelligence (AI) and cognitive science. It emphasizes the psychological, social, and philosophical implications of AI and, by means of an extended project to design an Automated Tourist Guide, makes the connection between the details of an AI programming language and the 'magic' of artificial intelligence programs, which converse in English, solve problems, and offer reasoned advice.

The book covers computer simulation of human activities, such as problem solving and natural language understanding; computer vision; AI tools and techniques; an introduction to AI programming; symbolic and neural network models of cognition; the nature of mind and intelligence; and the social implications of AI and cognitive science.

Each chapter will, in general, present a particular AI topic, with sections on the background to the topic, methods, and applications. These do not assume any previous knowledge of a computer language, and are intended for the reader who wants to gain an understanding of the field without plunging into programming. The foreword and chapter 1 offer an overview of artificial intelligence and cognitive science. The fundamental AI techniques of pattern matching, knowledge representation, and search are covered in chapters 2–4, and these chapters need to be read thoroughly in order to get a grounding in the subject. Chapters 5–9 deal with applications of AI and the techniques of reasoning with stored knowledge. They can be read out of order, or to different depths, although the programming appendix for a chapter will use terms introduced in earlier ones. Chapter 10 is a fairly self-contained discussion of AI and the philosophy of mind. Chapter 11 speculates on the future of AI and its social implications.

Most of the chapters contain an appendix which presents the topic in terms of an AI programming language. The language we have chosen, POP-11, is not the most widely used one for AI (although it is rapidly growing in popularity), but it is a language both for beginners and for advanced research. These aims are not contradictory, since both beginners and advanced programmers need tools that are well designed, and that reveal the structure of the problem at hand. POP-11 is particularly suitable as a way of describing programs on paper, since its appearance is similar to the popular teaching language PASCAL, and it encourages clear, well-structured programs. You do not need to have access to a computer running POP-11; the programs that we use as examples can be followed on the printed page.

Wherever possible we have used plain language and avoided technical terms unless they are an essential part of the vocabulary of AI. Such words are printed in boldface and included in the glossary.

Although this book is intended to give you a good feel for the issues and practicalities of AI and cognitive science, it does not attempt a full coverage of the subject. Nor in general does it go into the details of programming or computer science. It should be seen as the text for a course on "Foundations of AI and Cognitive Science" or as a preliminary to more technical texts such as those by Charniak and McDermott (1985), Rich (1983), and Winston (1984) for AI, and Stillings (1987) for cognitive science.

The book arose from a 10-week course for first-year arts undergraduates at Sussex University, also called "Computers and Thought." Most of the students have no previous experience of computing, and many of them are deeply suspicious of what they see as attempts to replace people with computers. As well as introducing them to the tools and methods of AI we have tried to show that, by building models of the mind, we can uncover the fascinating range and detail of human cognition; by attempting, and failing, to build thinking machines we gain respect for living, thinking beings.

One person took prime responsibility for each of the chapters (apart from chapter 5, which was jointly written by David Hogg and Chris Hutchison). Mike Sharples wrote the main draft of chapters 1, 8, and 11. David Young wrote the main draft of chapters 2 and 4. Steve Torrance wrote the main draft of chapters 3 and 10. Chris Hutchison wrote the main draft of chapters 6 and 7. David Hogg wrote the main draft of chapter 9. After comments from independent reviewers, the manuscript was reworked by the original authors and by Mike Sharples, to create a consistent style and to tidy up cross references. We believe we now have the best of both worlds: one coherent textbook, built from chapters written by specialists on their own subject areas.

Over the years the "Computers and Thought" course has been revised and polished by many people. We have drawn on course material produced by Aaron

Sloman, Steve Hardy, Benedict du Boulay, and Tom Khabaza. We should like to thank them and the other staff, the POPLOG team who developed POP-11, the reviewers who provided detailed comments on the chapters, and the students at Sussex University who have provided the fertile soil for this book. Particularly, we want to thank Robert Bolick of The MIT Press for championing the book and for his patient and helpful advice, Harry Stanton of MIT Press/Bradford Books for easing it through all the stages of production, and Nick Franz for spending many hours with LaTeX turning a manuscript into a book.

The coloring of the image on the cover is by the British artist Harold Cohen. The drawing is the work of Cohen's computer program, AARON. AARON is an intelligent computer-based program, now in its fifteenth year of continuous development, and the only program currently in existence capable of the autonomous generation of original works of art. Harold Cohen's goal in writing AARON has been to understand how human beings make and read images, not to simulate existing works of art. Cohen and AARON have exhibited together in art museums and science centers in New York, London, Tokyo, Amsterdam, Toronto, San Francisco, Boston and many other major cities.

Foreword: A Personal View of Artificial Intelligence

Introduction

There are many books, newspaper reports, and conferences providing information and making claims about artificial intelligence (AI) and its lusty infant, the field of expert systems. Reactions range from one lunatic view that all our intellectual capabilities will be exceeded by computers in a few years' time to the slightly more defensible opposite extreme view that computers are merely lumps of machinery that simply do what they are programmed to do and therefore cannot conceivably emulate human thought, creativity, or feeling. As an antidote for these extremes, I shall try to sketch a sane middle-of-the-road view.

In the long-term, AI will have enormously important consequences for science and engineering and our view of what we are. But it would be rash to speculate in detail about this. In the short-to-medium term there are extremely difficult problems. The main initial practical impact of AI will arise not so much from intelligent machines as from the use of AI techniques to build 'intelligence amplifiers' for human beings. Even if machines have not advanced enough to be capable of designing other complex machines, discovering new concepts and theories, understanding speech at cocktail parties, and making all our important economic, political, and military decisions for us, AI systems may nevertheless be able to help people to learn, plan, take decisions, solve problems, absorb information, find information, design things, communicate with one another, or even just explore ideas when confronted with a new problem.

Besides helping human thought processes, AI languages, development tools, and techniques can also be used for improving and extending existing types of automation, for instance: cataloguing, checking computer programs, checking consistency of data, checking plans or designs, formatting documents, analyzing images, and many kinds of monitoring and controlling activities.

What Then Is AI?

Some people give AI a very narrow definition as an applied sub-field of computer science. I prefer a definition that reflects the range of work reported at AI conferences, in AI journals, and the interests and activities of some of the leading practitioners, including founders of the subject. From this viewpoint AI is a very general investigation of the nature of intelligence and the principles and mechanisms required for understanding or replicating it. Like all scientific disciplines it has three main types of goals: theoretical, empirical, and practical.

Goals of AI: The Trinity of Science

The long-term goals of AI include finding out what the world is like, understanding it, and changing it, or, in other words,

a. empirical study and modelling of existing intelligent systems (mainly human beings);

b. theoretical analysis and exploration of possible intelligent systems and possible mechanisms and representations usable by such systems; and

c. solving practical problems in the light of (a) and (b), namely:

   c.1. attempting to deal with problems of existing intelligent systems (e.g., problems of human learning or emotional difficulties) and

   c.2. designing useful new intelligent or semi intelligent machines.

Some people restrict the term 'artificial intelligence' to a subset of this wide-ranging discipline. For example, those who think of it as essentially a branch of engineering restrict it to (c.2). This does not do justice to the full range of work done in the name of AI.

In any case, it is folly to try to produce engineering solutions without either studying general underlying principles or investigating the existing intelligent systems on which the new machines are to be modelled or with which they will have to interact. Trying to build intelligent systems without trying to understand general principles would be like trying to build an aeroplane without understanding principles of mechanics or aerodynamics. Trying to build them without studying how people or other animals work would be like trying to build machines without ever studying the properties of any naturally occurring object.

The need to study general principles of thought, and the ways in which human beings perceive, think, understand language, etc., means that AI work has to be done in close collaboration with work in psychology, linguistics, and even philosophy, the discipline that examines some of the most general presuppositions of our thought and language. The term 'cognitive science' can also be used to cover the full range of goals specified above, though it too is ambiguous, and some of its more narrow-minded practitioners tend to restrict it to (a) and (c.1).

But What Is Intelligence? — Three Key Features

The goals of AI have been defined in terms of the notion of intelligence. I do not pretend to be able to offer a definition of 'intelligence'. However, most, if not all, of the important work in AI arises out of the attempt to understand three key characteristics of the kind of intelligence found in people and, to different degrees, other animals. The features are intentionality, flexibility, and productive laziness.

Intentionality

This is the ability to have internal states that refer to or are *about* entities or situations more or less remote in space or time, or even non-existent or wholly abstract things.

So intentional states include contemplating clouds, dreaming you are a duke, exploring equations, pondering a possible action, seeing a snake, or wanting to win someone's favours. These are all cases of awareness or consciousness of something, including hypothetical or impossible objects or situations. A sophisticated mind may also have thoughts or desires about its own state — various forms of *self*-consciousness are also cases of intentionality.

All intentional states seem to require the existence of some kind of *representation* of the content of the state: some representation of whatever is believed, perceived, desired, imagined, etc. A major theme in AI is therefore investigation of different kinds of representations and their implementation and uses. This is a very tricky topic, since there are many different kinds of representational forms: sentences, logical symbols, computer databases, maps, diagrams, arrays, images, etc. It is very likely that there are still important forms of representation waiting to be discovered.

Flexibility

This has to do with the breadth and variety of intentional contents, i.e., the variety of types of things intentional states can refer to, for instance, the variety of types of goals, objects, problems, plans, actions, environments etc., with which an individual can cope, including the ability to deal with new situations using old resources combined and transformed in new ways.

Flexibility in this sense is required for understanding a sentence you have never heard before, seeing a familiar object from a new point of view, coping with an old problem in a new situation, and dealing with unexpected obstacles to a plan. A kind of flexibility important in human intelligence involves the ability to raise a wide range of questions.

A desirable kind of flexibility often missing in computer programs is 'graceful degradation'. Often if the input to a computer deviates at all from what is expected, the result is simply an error message and abort. Graceful degradation, on the other hand, would imply being able to try to cope with the unexpected by reinterpreting it, or modifying strategies, or asking for help, or monitoring actions more carefully. Instead of total failure, degradation might include taking longer to solve a problem, reducing the accuracy of solution, reducing the frequency of success, etc.

One of the factors determining the degree of flexibility will be the range of representations available. A system that can merely represent things using a vector of numerical measures, for example, will have a narrower range of possible intentional states than a system that can build linguistic descriptions of unlimited complexity, like:

the man
the old man
the old man in the corner
the old man sitting on a chair in the corner
the sad old man sitting on a chair with a broken leg in the corner
etc.

so flexible control systems of the future will have to go far beyond using numerical measures, and will have to be able to represent goals or functions, and relationship between structures, resources, processes, constraints, and so on.

Productive Laziness

It is not enough to achieve results: intelligence is partly a matter of *how* they are achieved. Productive laziness involves avoiding unnecessary work.

A chess champion who wins by working through all the possible sequences of moves several steps ahead and choosing the best one is not as intelligent as the player who avoids examining so many cases by noticing that the pieces form a pattern which points directly to the best move.

Why is laziness important? Given any solvable task for which a finite solution is recognizable, it is possible in principle to find a solution by enumerating all possible actions (or all possible computer programs) and checking them exhaustively until the right one turns up. In practice this is useless because the set of possibilities is too great.

This is called a 'combinatorial explosion'. Any construction involving many choices from a set of options has a potentially huge array of possible constructs to choose from. If you have 4 choices each with 2 options, the total set of options is 16. If you have 20 choices each with 6 options, the total shoots up to 3,656,158,440,062,976. Clearly exhaustive enumeration is not a general solution, so lazy shortcuts have to be found.

For example, a magic square is an array of numbers all of whose rows, columns. and diagonals add up to the same total. Here is a 3 by 3 magic square made of the digits 1–9:

```
6 7 2
1 5 9
8 3 4
```

If you try to construct a 3 by 3 magic square by trying all possible ways of assigning the 9 numbers to the locations in the square, then there are 362,880 possible combinations. Trying them all would not be intelligent. A more sensible procedure would involve testing partial combinations to see whether they can possibly be extended satisfactorily, and, if not, rejecting at one blow all the combinations with that initial sequence. It is also sensible to look for symmetries in the problem. Having found that you cannot have the number 5 in the top left corner, reject all combinations that involve 5 in any corner.

Yet more subtle arguments can be used to prune the possibilities drastically. For example, since eight different triples with the same total are needed (one for each row, one for each column and the two diagonals), it is easy to show that large and small numbers must be spread evenly over the triples, and that they must in fact add up to 15. So the central number has to be in four different triples adding up to 15, the corner numbers in three triples each, and the mid-side numbers in two each. For each number we can work out how many different triples it can occur in, and this immediately restricts the locations to which they can be assigned. For example, 1 and 9 must go into locations in the middle of a side, and the only candidate for the central square is 5. In fact, a high-level symmetry shows that you need bother to do this analysis only for the numbers 1–4. You can then construct the square in a few moves, without any trial and error. What about a 2 by 2 magic square containing the numbers 1, 2, 3, and 4? Think about it!

These examples show that the ability to detect shortcuts requires the ability to *describe* the symmetries, relationships, and implications in the structure of the task. It also requires the ability to *notice* them and perceive their relevance, even though they are not mentioned in the statement of the task. This kind of productive laziness therefore depends on intentionality and flexibility, but motivates their application. Discovering relevant relationships not mentioned in the task specification (e.g., "Location X occurs in fewer triples than location Y") requires the use of a generative conceptual system and notation (i.e., one that enables novel descriptions to be formulated). Being lazy in this way is often harder than doing the stupid exhaustive search. But it may be very much faster. This points to a need for an analysis of the notion of intellectual difficulty.

Productive laziness often means applying previously acquired knowledge about the problem or some general class of problems. So it requires learning: the ability to form new concepts and to acquire and store new knowledge for future applications. Sometimes it involves creating a new form of representation, as has happened often in the history of science and mathematics.

Laziness motivates a desire for generality — finding one solution for a wide range of cases can save the effort of generating new solutions. This is one of the major motivations for all kinds of scientific research. It can also lead to errors of over-generalization, prejudice, and the like. A more complete survey would discuss the differences between avoiding mental work (saving computational resources) and avoiding physical work.

An (Overly) Simple Design for an Intelligent System

Here is a simple set of components for an intelligent system.

Perceptual mechanisms These mechanisms analyze and interpret information taken in by the "senses" and store the interpretations in a database.

A database of information This is not just as a store of facts, for a database can also store information about how to do things. It may include both particular facts provided by the senses and generalizations formed over a period of time.

Analysis and interpretation procedures These are procedures which examine the data provided by the senses, break them up into meaningful chunks, build descriptions, match the descriptions, etc. Analysis involves describing what is presented in the data. Interpretation involves describing something else, possibly lying behind the data, for instance, constructing a 3-D description on the basis of 2-D images, or inferring someone's intentions from his actions.

Reasoning procedures These use information in the database to derive further information which can also be stored in the database. If you know that Socrates is a man, and that all men are mortal, you can infer something new, namely, that Socrates is mortal.

A database of goals These just represent possible situations which it is intended should be made *actual.* There may also be policies, preferences, ideals, and the like.

Planning procedures These take a goal, and a database of information, and construct a plan which will achieve the goal, assuming the correctness of the information in the database.

Executive mechanisms and muscles or motors These translate plans into action.

Often the divisions will not be very clear. For instance, is 'this situation is painful' a fact or a goal concerned with the need to change the situation? This sort of system can be roughly represented by figure 1.

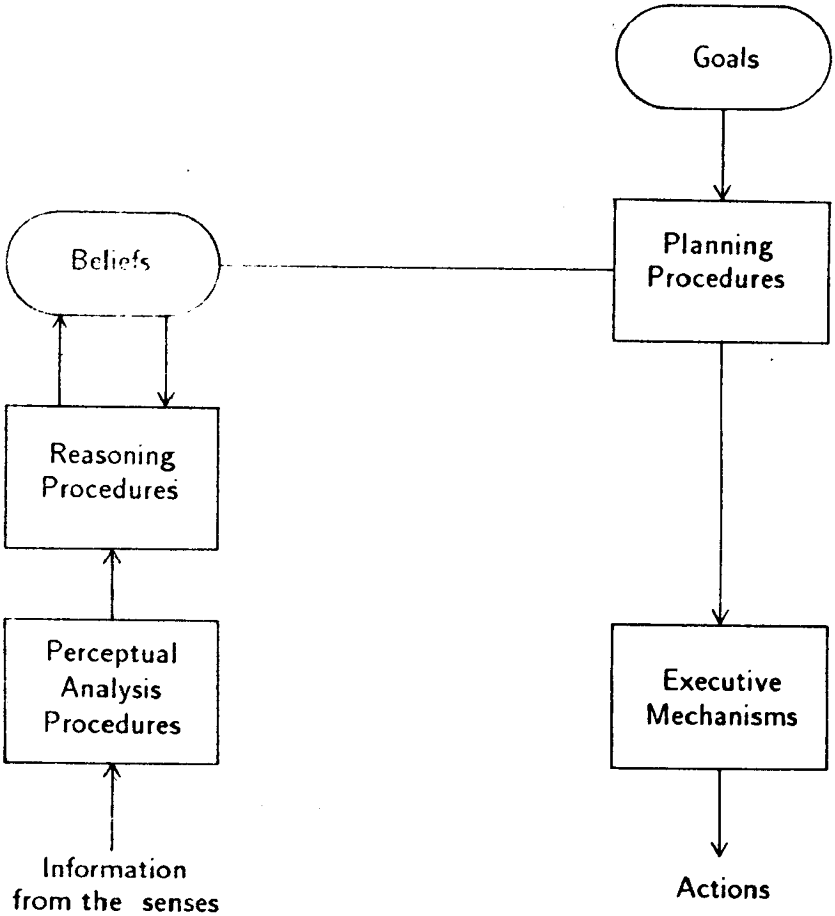

Figure 1 Sketch of a not very intelligent system.

Limitations of This Design

This sort of diagram conceals much hidden complexity. Each of the named sub-processes may have a range of internal structures and sub-processes, some relatively permanent, some very short term.

However, even this kind of complexity does not do justice to the kind of intelligence that we find in human beings and many animals. For example, there is a need for internal self-monitoring processes as well as external sensory processes. A richer set of connections may be needed between sub-processes. For example, planning may require reasoning, and perception may need to be influenced by beliefs, current goals, and current motor plans (see figure 1).

It is also necessary to be able to learn from experience, and that requires processes that do some kind of retrospective analysis of past successes and failures. The goals of an autonomous intelligent system are not static, but are generated dynamically in the light of new information and existing policies, preferences, and the like. There will also be conflicts between different sorts of goals that need to be resolved. Thus 'goal-generators' and 'goal-comparators' will be needed, and mechanisms for improving these in the light of experience.

Further complexities arise from the need to be able to deal with new information and new goals by interrupting, modifying, temporarily suspending, or aborting current processes. I believe that these are the kinds of requirements that explain some kinds of emotional states in human beings, and we can expect similar states in intelligent machines.

Whether or not the design sketched above is accurate, ideas developed in exploring such designs may prove to be essential for developing correct theories about how the mind works. This may be so even if the human mind is embodied in a physical system whose basic mechanisms are very different from a modern digital computer.

'Noncognitive' States and Processes

One of the standard objections to AI is that although it may say something useful about *cognitive* processes, such as perception, inference, and planning, it says nothing about other aspects of mind, such as motivation and emotions. In particular, AI programs tend to be given a single 'top-level' goal, and everything they do is subservient to this, whereas people have a large number of different wishes, likes,