THE APIC SERIES · 32 ·

KNOWLEDGE REPRESENTATION

An Approach to Artificial Intelligence

T.J.M. BENCH-CAPON

The A.P.I.C. Series
No 32

Knowledge Representation

An Approach to Artificial Intelligence

T.J.M. BENCH-CAPON
Department of Computer Science,
University of Liverpool, England

ACADEMIC PRESS

Harcourt Brace Jovanovich, Publishers
London San Diego New York Boston
Sydney Tokyo Toronto

ACADEMIC PRESS LIMITED
24/28 Oval Road,
London NW1 7DX

United States Edition published by
ACADEMIC PRESS INC.
San Diego, CA 92101

This book is printed on acid-free paper

British Library Cataloguing in Publication Data

Is available

ISBN 0-12-086440-1

Typeset by Colset, Pte. Ltd., Singapore
Printed in Great Britain by
Galliard (Printers) Ltd, Great Yarmouth, Norfolk

Knowledge Representation

An Approach to Artificial Intelligence

This is volume 32 in the A.P.I.C. Series

General Editors: M.J.R. Shave *and* I.C. Wand

A complete list of titles in this series appears at the end of this volume

Preface

This book is about knowledge representation. Knowledge representation has emerged as one of the fundamental topics in the area of computer science popularly known as *artificial intelligence* (AI). The reason for this is simple: the basic idea is that intelligence, whatever else it involves, does at least involve knowing things, and exploiting them so as to respond appropriately to a given situation. Therefore it seems reasonable to suppose that if we wish to make an intelligent computer system we must have a way of getting it to know things, and that involves finding a way of representing the things we wish it to know so that they can be encoded within the computer system. Thus the intention of the book is to provide an introduction to the field of artificial intelligence, via an examination of issues concerning knowledge representation.

Whilst a variety of paradigms that have been used in artificial intelligence will be considered, there will be some focus on those which might be termed *rule based* at the expense of those which might be termed *structured object* representations. This emphasis is in part due to my personal predilections, but none the less I have tried to give reasonable coverage also to competing ideas. The first four chapters will be concerned with general issues, and will provide background to the rest of the book, which will comprise a detailed consideration of three of the major knowledge representation paradigms which have come out of AI research, and thus represent much of the cumulative wisdom on the matter.

In the first chapter, we will consider what is meant by "artificial intelligence", and the related terms like *knowledge-based systems* and *expert systems*, so that we can be clear about our motivation, and about exactly what we are trying to achieve with such programs. The second chapter gives a general consideration of what knowledge representation is, and what criteria we can use to assess a particular form of knowledge representation. The third chapter outlines some of the basic concepts and terminology of formal logic, which is essential as a means of describing and distinguishing between the various computer-oriented paradigms discussed later. The fourth chapter gives a description of the key topics relating to *search*, which is historically important in AI, and which is also crucial to the understanding of a range of knowledge representation issues.

Chapters 5, 6 and 7 consider the three leading representation paradigms—production rules, structured object representations and predicate calculus, —so that we can see the style of systems they produce, and consider their various strengths and weaknesses. To illustrate some of the use that has been made of these paradigms the eighth chapter examines PROLOG, which can be seen as a practical instantiation of many of the ideas deriving from the predicate logic paradigm, and the ninth chapter discusses *expert systems*, which represent an area of application which has made extensive use both of production rules and structured object representations.

In the last chapter a number of issues relating to the use and limitations of the paradigms will be discussed, so as to indicate some of the unsolved problems of knowledge representation.

The book originated in a course of lectures given by me to third-year undergraduates in computer science at the University of Liverpool and is directed primarily at undergraduates embarking on a first course in artificial intelligence. Whilst I have tried to presuppose as little knowledge as possible, a reader will need some familiarity with the concepts of computer science, and would be helped by such grounding in formal logic as might be given by an introductory course in the subject.

I would like to take this opportunity to thank all those with whom I have discussed knowledge representation issues: particularly Marek Sergot, Robert Kowalski and my former colleagues at Imperial College, Justin Forder and my former colleagues on the Alvey–DHSS Demonstrator project, and my present colleagues at the University of Liverpool. Thanks are due also to those who have read and commented on earlier drafts, especially my wife, Priscilla, who contributed many valuable comments which improved the grammar, spelling and comprehensibility. Needless to say, remaining errors and unsupported prejudices are my own.

Contents

1

Introduction to Artificial Intelligence and Knowledge-Based Systems

1.1. Artificial and intelligent

The term "artificial intelligence"—almost always abbreviated to "AI"—is convenient both because it gives an intuitive characterisation of what this area of computer science is about, and because it conveys an air of excitement which stimulates enthusiasm. But in other ways it is something of a liability, in that it often arouses unjustifiable expectations, and because any precise definition is elusive. In this section I want to examine some of the various things that people have meant by the term. The purpose is not to give a tight definition, but rather to create an awareness of the issues involved. So let us begin by asking what would be expected from an artificially intelligent system. Clearly, if tautologically, we would expect it to be both artificial and intelligent. The "artificial" part gives little problem. A thing is artificial in the sense under discussion simply if it is made by man, although perhaps there are also some connotations of "being made in imitation"—the sense in which "artificial" is opposed to "real", as in "an artificial smile"—so that we could see an AI system as one which attempts to imitate natural intelligence. But in the context of computer science we may well be satisfied with the idea of artificial as simply being a matter of being embodied in a computer system—which is clearly made by man, and which may be intended to imitate human behaviour. Thus an AI system can, for our purposes, be taken to mean an intelligent computer system.

Now some philosophers have argued that the idea of an intelligent computer system is simply absurd, and that it can be shown that no computer system could be intelligent. The champion of this stance in the early years of AI was Hubert Dreyfus [1], while currently it is John Searle [2]. Their arguments turn on their views on philosophy of mind and are proper to philosophy rather than computer science. Searle's main argument derives from a "thought experiment" in which a person is shut in a room and given a set of rules, written in English, to manipulate Chinese characters. He is then presented with sequences of Chinese characters and follows the rules so as to produce further sequences of Chinese characters. To those outside the room it might appear that they were asking questions in Chinese, and the man in the room was replying in Chinese and so they might conclude that he understood Chinese, whereas it is clear that he does not: he is merely following rules in an unthinking fashion. Searle claims that there is no understanding of Chinese here, and no intelligence in the replies, and then contends that the same would be true for any computer system embodying a formal program. What is missing is understanding. Stated in its sharpest form, Searle's claim is that even if the behaviour of a machine acting in accordance with a formal program was indistinguishable from that of a person, we should not ascribe intelligence to it if we knew that it was a machine, although we might mistakenly do so if we were unaware that it was a machine. The variety of responses to Searle [2] indicate, however, that workers in AI would count such a system as intelligent. The disagreement really turns on what is meant by "intelligence": for Searle it is inextricably bound up with the notion of "intentionality", which goes beyond observable behaviour. This is not the place to discuss intentionality, which is a rather unclear notion in philosophy; rather we should note that the possession of intelligence is not a straightforward thing, and that should we wish to characterise our work as the building of intelligent computer systems, we need to be able to say what we mean by this, in the context of computer science. Even if we were to accept the arguments of these philosophers it would show no more than that "artificial intelligence" is something of a misnomer, because there would remain the body of work done under the AI banner, and this work would remain worthy of study. What we need to do, therefore, is find what is meant by "intelligent" in this specialised context.

1.1.1. Popular views of intelligent machines

It is, however, worth noting in passing that the popular imagination has never experienced difficulty in attributing intelligence to machines. In the coffee houses of eighteenth century France a gambler made a fair sum of money by inducing people to play chess against what he claimed was a chess playing automaton, known as the Turk. That it was a fraud consisting of a dwarf in a box was discovered by inspection rather than reflection on the absurdity of an intelligent machine. (In those days it was considered necessary to possess intelligence to play chess.)

Similarly, when Babbage developed his Analytical Engine, in the early nineteenth century, so eager were people to ascribe intelligence to it that disclaimers were felt necessary. For example we find Ada, Lady Lovelace, Babbage's associate, writing

> [It] has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis, but it has no power of originating analytical relations or truths. Its province is to assist us in making available what we are already acquainted with. [3]

If we consider the public conception of computers in the twentieth century we find a similar phenomenon. The public press often refers to computers as "electronic brains" and people happily watch films like 2001 in which the computer HAL is accepted as more or less the only intelligent being around. Computer scientists, in contrast, used to be at pains to emphasise that the computers are really pretty dumb, relying on pure speed of computation, and following the orders of their programmers to the letter, no matter how silly the results. The issue of bills for £0.00 is a recurring source of entertainment for the popular media as an example of the alleged stupidity of computers. On this view computers are seen as "fast idiots" rather than "electronic brains".

The growth of AI has confused these traditional terms of the debate because we find some computer scientists wishing to ascribe intelligence to their systems, while, of course, retaining the conventional computer science view for conventional systems. The whole problem really turns on the definition of "intelligence". I would contend that the underlying definitions used by philosophers like Dreyfus and Searle, which derive from particular philosophies of mind, are not really in correspondence with the popular notion. But neither is the extreme opposite used by the AI scientist who said that his thermostat was intelligent because it knew when the room was hot and turned the temperature down. There does seem to be a need to point to some requirement for the use of what we might, at the risk of begging the question, term "higher mental processes" in the task before we wish to ascribe intelligence to its performance. Behaviour of the sort that is no more than a pre-determined response to a stimulus, as is the case with the thermostat, is not enough. What we want to see, and what those striving to build AI systems want to see, is the computer "originating" something. We want to see it telling us something we did not know and would not have thought of, or worked out from the information supplied to the system, before we feel that we should ascribe intelligence to it.

1.1.2. Human versus machine intelligence

In discussing the notion of intelligence as applied to computer systems, however, we should be aware that what counts as an intelligent computer system is not the same as what we would count as intelligent for a human being. In particular, if we see a child capable of accurately performing complex pieces of arithmetic, we should be likely to think him clever, whereas a child who was unable to recognise his family and friends by looking at them would be thought very dull indeed. Quite the reverse is true of computers: the ability to do arithmetic is so taken for granted that this ability counts for nothing in assessing the claims to intelligence of a program, whereas the recognition of faces, and even simple language understanding, is considered to be firmly within the field of AI. This view makes intelligence relative to a norm: every sort of thing, be it a person or a machine, has a set of competences which is expected of it, and only where it goes beyond these do we start to talk of intelligence.

This is the kind of consideration that has led to some popular definitions of AI; such as that offered by Rich in her book *Artificial Intelligence*, that

> AI is the study of how to make computers do things, at which, at the moment, people are better. [4]

On this view, at present people are better at recognising faces, while computers are better at arithmetic; so the former is AI and the latter is not.

1.1.3. Turing test

The definition given by Rich is still pretty vague, however, so if we are to attempt to decide in any objective sense whether or not a system is intelligent, we must propose some precise criterion for the presence or absence of intelligence. The most frequently cited test of this sort is the *Turing test* first proposed by Alan Turing [5]. In this test a person converses with another person and the putatively intelligent system over two teletypes. He asks a series of questions and is supposed to decide, on the basis of the answers, which set of responses come from the person and which from the computer system. If he is unable to decide, then the system passes the Turing test and may be considered intelligent. The test was based on a party game where a person conversed through another person with two people of different sex and was supposed to determine which was the man and which was the woman from their responses.

Despite its appeal, the Turing test runs into two kinds of difficulty. First it sets an impossibly high standard. If the user is permitted to ask any questions whatsoever, and Turing's original statement of the test makes it clear that this is the intention, the test is unlikely to be achieved in the next couple of generations. All successful AI programs, even those whose success is modest, have achieved such success as they have achieved only by focusing on a limited domain. Even to parse an indefinite range of questions is well beyond our current capabilities, let alone to use the sophisticated answering strategies that Turing seems to have in mind. As an example of this game-playing strategy, in his original article an illustrative test asks a complicated multiplication question and the machine not only pauses before answering, but gets it wrong as well.

The second kind of difficulty is that the test is, if applied with sufficient generosity to allow for the possibility of success, too easy. For programs that have been written have managed to pass, admittedly under propitious conditions, the Turing test, even though no one would even begin to claim that the programs in question manifested intelligence in any real sense. ELIZA and PARRY are the classic examples [6]. The problem here is that the user of the system attempts to supply meaning to the output, and can place an intelligent construction on behaviour which is in fact purely mechanical. Thus the Turing test, though an interesting idea and often used as a discussion point, fails to provide us with a practical touchstone.

1.1.4. Typical AI applications

Given the inconclusive nature of the above discussion we might well feel that it would be better to define AI not by the attempt to create intelligence in a computer system, but rather by reference to the kinds of activity that are well accepted as being within the AI domain. This way we do not have to define "intelligence" at all. On this view, we do not take the constituent terms of artificial intelligence seriously at all, but use it as an appealing label for a group of applications. Thus we could point to a long list of applications that would be accepted as within this area of computer science. We could suggest game playing, problem solving, solution of IQ tests, language translation, theorem proving, natural language understanding, planning, vision, speech recognition, robotics and expert systems as a reasonable, although not exhaustive, list. This approach is consistent with the one taken by Rich when she gave the definition cited above. Her definition has the advantage of picking out an identifiable body of work, much as given above, but it suffers from two disadvantages. First, and most obvious, it is a definition which will be constantly shifting. Any successful AI program will have the effect, by virtue of its very success, of ceasing to be an AI program. Perhaps this does not really matter, given the slow rate of progress in AI, but it seems to make AI a fairly unrewarding field in which to work. It is really rather too close to the cynic's definition of AI, that "AI is what we don't know how to do yet", for complete comfort.

A deeper criticism, which applies to any attempt to characterise AI by reference to a list of applications, derives from the suggestion that "it ain't what you do it's the way that you do it". Chess programs are a case in point here. Game playing in general, and chess in particular, have held attractions for workers in AI, perhaps because chess is seen as the intellectual game *par excellence*. Because of limitations in the power and capacity of early