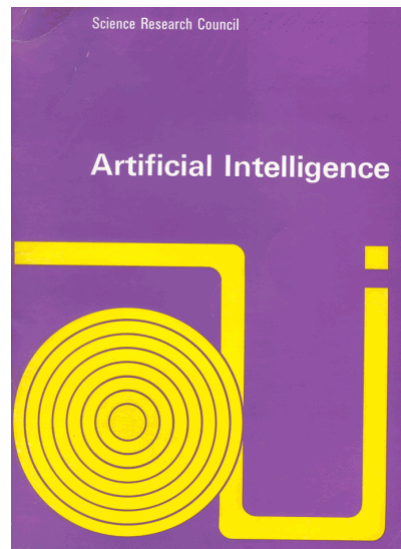


Lighthill Report: Artificial Intelligence: a paper symposium



Lighthill Report, 1973

[Full Size Image](#)

- [Preface](#)
- [Contents](#)
- [Part I: Artificial Intelligence: A General Survey](#) by Professor Sir James Lighthill, FRS
- [Part II: Some Comments on the Lighthill Report and on Artificial Intelligence](#) by Professor N S Sutherland
- Part III: Comments on the Lighthill Report and the Sutherland Reply
- [By Dr R M Needham](#)
- [By Professor H C Longuet-Higgins, FRS](#)
- [By Professor D Michie](#)

Source:

http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm

Overview

Lighthill's report was commissioned by the Science Research Council (SRC) to give an unbiased view of the state of AI research primarily in the UK in 1973. The two main research groups were at Sussex and Edinburgh. There was pressure from Edinburgh to buy a US machine, the Digital Equipment Corporation DEC10 which was used by most US researchers. AI research was funded by the Engineering Board of SRC as part of its Computer Science funding. The Lighthill Report was published early in 1973. Although it supported AI research related to automation and to computer simulation of neurophysiological and psychological processes, it was highly critical of basic research in the foundation areas such as robotics and language processing. Lighthill's report provoked a massive loss of confidence in AI by the academic establishment in the UK including the funding body. It persisted for almost a decade.

AI research continued but the next attempt to mount a major activity in the area did not come until the September 1982 Research Area Review Meeting on Intelligent Knowledge-Based Systems. The findings of which were accepted by SERC (Science and Engineering Research Council, a change of name) and became the IKBS part of the Alvey Programme.

A video of a BBC TV Debate - June 1973 - Lighthill Controversy - at the Royal Institution with Professor Sir James Lighthill, Professor Donald Michie, Professor Richard Gregory and Professor John McCarthy concerning the findings is available at <http://www.aiai.ed.ac.uk/events/lighthill1973/>.

Sir James Lighthill died on the 17 July 1998.

Stuart Sutherland died on the 8 November 1998.

Roger Needham died on the 28 February 2003.

Hugh Christopher Longuet-Higgins died on 27 March 27 2004.

Donald Michie died in a car accident on 7 July 2007.

Artificial Intelligence: A General Survey

Professor Sir James Lighthill FRS

Part I Artificial Intelligence

A general survey by Sir James Lighthill

FRS Lucasian Professor of Applied Mathematics, Cambridge University. July 1972.

1 Introduction

The Science Research Council has been receiving an increasing number of applications for research support in the rather broad field with mathematical, engineering and biological aspects which often goes under the general description Artificial Intelligence (AI). The research support applied for is sufficient in volume, and in variety of discipline involved, to demand that a general view of the field be taken by the Council itself. In forming such a view the Council has available to it a great deal of specialist information through its structure of Boards and Committees; particularly from the Engineering Board and its Computing Science Committee and from the Science Board and its Biological Sciences Committee. These include specialised reports on the contribution of AI to practical aims on the one hand and to basic neurobiology on the other, as well as a large volume of detailed recommendations on grant applications.

To supplement the important mass of specialist and detailed information available to the Science Research Council, its Chairman decided to commission an independent report by someone outside the AI field but with substantial general experience of research work in multidisciplinary fields including fields with mathematical, engineering and biological aspects. I undertook to make such an independent report, on the understanding that it would simply describe how AI appears to a lay person after two months spent looking through the literature of the subject and discussing it orally and by letter with a variety of workers in the field and in closely related areas of research. Such a personal view of the subject might be helpful to other lay persons such as Council members in the process of preparing to study specialist reports and recommendations and working towards detailed policy formation and decision taking.

The report which follows must certainly not be viewed as more than such a highly personal view of the AI field. It owes much to the study of published work and of private written communications and spoken comments by numerous individuals, including the following:

J. Annett, H. G. Barrow, S. Brenner, D. E. Broadbent, R. A. Brooker, O. P. Buneman, R. M. Burstall, A. D. B. Clarke, M. B. Clowes, A. H. Cook, D. C. Cooper, J. E. Doran, J. F. Duke, E. W. Elcock, I. J. Good, C. C. Green, R. L. Gregory, P. J. Hayes, A. L. Hodgkin, J. N. Holmes, J. A. M. Howe, D. H. Hubel, S. Isard, H. Kay, T. Kilburn, J. Kulikowski, D. N. L. Levy, H. C. Longuet-Higgins, D. M. MacKay, D. Marr, J. McCarthy, B. Meltzer, D. Michie, M. Minsky, D. Mollison, E. Moore, J. S. Moore, R. M. Needham, N. J. Nilsson, C. Oldfield, J. V. Oldfield, I. Pohl, R. J. Popplestone, B. Raphael, J. A. Robinson, C. Strachey, N. S. Sutherland, M. M. Swann, H. P. F. Swinerton-Dyer, D. Wilshaw, T. Winograd.

The author is grateful for the large amount of help and advice readily given in reply to his many requests. He must emphasize, however, that none but himself is responsible for the opinions expressed in this report. They represent merely the broad overall view of the subject which he reached after such limited studies as he was able to make in the course of two months.

Readers might possibly have expected that the report would include a summary, but the author decided against this partly because considerable material is summarised already in

almost every paragraph. Furthermore, he believes that this kind of report can be valuable only to those who read it all, and for this reason preferred to avoid attempting a condensation.

2 The ABC of the subject

There is a general consensus about which main areas of research are to be grouped within the broad field of AI. This section lists briefly these main areas and divides them further into three categories, A, B and C according to the long-term motivations for the three different types of work.

Here, categories A and C have clearly distinct motivations: each has a well defined general direction of its intended objectives, but the two directions are quite different. In both these categories a certain amount of rather respectable progress has been made during the subject's twenty-five years of life (which may be taken as beginning with Turing's 1947 article [Intelligent Machinery](#)), although expectations have as we shall see in [section 3](#) been frequently disappointed. During the same period a further category B of researches has been pursued: a [bridge](#) category where aims and objectives are much harder to discern but which leans heavily on ideas from both A and C and conversely seeks to influence them. Research in category B, if acceptable arguments for doing it can be agreed, works by its interdependence with studies in categories A and C to give unity and coherence to the whole field of AI studies. There is, however, a widespread feeling ([Section 3](#)) that progress in this bridge category B has been even more disappointing, both as regards the work actually done and as regards the establishment of good reasons for doing such work and thus for creating any unified discipline spanning categories A and C.

Category A

Here, letter A stands for Advanced Automation: the clear objective of this category of work being to replace human beings by machines for specific purposes, which may be industrial or military on the one hand, and mathematical or scientific on the other. The work looks beyond automation of the type that is widely adopted at present in control engineering and data processing, and aims to make a far fuller use of the general-purpose digital computer's logical (as opposed to arithmetical) potentialities. Nevertheless it must be looked at as a natural extension of previous work on the automation of human activities, and be judged by essentially the same criteria.

Industrially important purposes include, for example, machine recognition of printed or typewritten characters (an area where good progress has been made) and of handwritten characters (incomparably more difficult), as well as a much wider range of pattern-recognition activities. The auditory equivalent to this visual area is speech recognition and synthesis. There are great economic incentives for work in machine recognition of speech, as well as in machine translation between languages, although progress in both has so far been very disappointing.

A further industrially important aim is to go beyond the automation of component design and manufacture, towards automation of design and assembly of whole products. It is argued that the complex spatial relationships involved in assembly processes put them far beyond the scope of conventional control engineering and require a much more advanced logical structure in the controlling software. Similar arguments may apply to problems of improving packing ratios in parcel containerisation.

The level of automation that can be called advanced has to be placed higher in the military field with its remarkable achievements both in cryptography and in guided missiles. A modern missile's capability to move in response to its own perception of its target against a noise background is highly reminiscent of the way in which a predator uses its complex

central nervous system to home on to its prey. Beyond this the military have an incentive, however, to build less specialised devices that might be programmed to perform in hostile environments a far wider range of actions in response to information from organs of perception. Space exploration and, perhaps, some parts of industry (including fire-fighting) may look for a similar hostile-environment capability.

In the meantime, the application of digital computers in mathematical work has gone beyond the mere organisation of numerical calculations on a large-scale and includes for example some very effective programmes for massive manipulations in algebra and analysis. Category A looks well beyond these, however, to the automation of problems of logical deduction including theorem proving, and still further to the automation of inductive generalisation and [analogy spotting](#).

In scientific applications, there is a similar look beyond conventional data processing to the problems involved in large-scale data banking and retrieval. The vast field of chemical compounds is one which has lent itself to ingenious and effective programs for data storage and retrieval and for the inference of chemical structure from mass-spectrometry and other data.

Information retrieval is, indeed, one of two dominant themes underlying all work in category A: this work is found to depend essentially on a [knowledge base](#) which the program causes to be stored in the computer, and the [file structure](#) of this knowledge base is of crucial importance in determining how data is accessed and used in the machine's operations. The other dominant theme is problem solving. This goes beyond mathematical theorem proving into the solution of numerous [common-sense problems](#) such as may arise in industrial and other applications. They can often be represented as problems of [transversing a graph](#), using [graph](#) in the specialised mathematical sense: an assemblage of points or nodes representing states of the system studied, some but not all pairs of nodes being linked by a line representing a permitted transition between states. Programs may be sought for solving problems in the sense of finding [optimal](#) (eg shortest) paths between remote nodes on such a graph.

Longer-term objectives in category A include that of combining a well structured knowledge base and an advanced problem solving capability to generate improved methods for industrial and economic planning and decision making, although admittedly there will always be serious difficulties in establishing that any particular program must necessarily have an acceptable output of plans and decisions! Another longer-term objective permeating all work in category A furthermore, is to incorporate into programs an increasingly greater capability of [learning](#), so as to reach improved levels of performance in response to experience with tasks already undertaken. Efficient modes of learning will, however, be seen in [section 3](#) to remain somewhat elusive.

To sum up, category A is concerned with Advanced Automation aimed at objectives such as written character recognition, pattern recognition, speech recognition and synthesis, machine translation, product design and assembly, container packing, exploration and action in hostile environments, theorem proving, inductive generalisation, analogy spotting, information storage and retrieval, analysis of chemical structures, problem solving, graph traversing, learning and decision making. In marked contrast to the diversity characteristic of all these numerous objectives, whether practical or scientific, is the much more unified fundamental-research aim of category C: computer-based studies related to the Central Nervous System (CNS) in man and animals.

Category C

Thus, letter C stands for Computer-based CNS research. In a lay person's report the extended term central nervous system is used in preference to the term brain which to a lay person may have subjective associations with the more conscious, or more consciously

[brainy](#), parts of the brain activity, as against the emotional parts or those associated with perception and movement. The co-ordination of perception and movement in animals generally is a particularly significant area of research which the lay person (unlike the biologist) might be tempted to forget if the word [brain](#) conjures up for him the specifically human aspects of brain activity.

Category C is concerned, then, with theoretical investigations related to neurobiology and to psychology. The word [theoretical](#) is used here to emphasize that we are concerned, not at all with the use of computers merely to process experimental data, but with their use to build models of CNS processes whose performance can in due course be compared with experimental data - a phrase carefully chosen to be uncontroversial as between critics of work in category C who argue that it takes insufficient account of existing data and some of the research workers who feel that the experimentalists will need the stimulus of revolutionary theoretical ideas to produce their best work.

It must be emphasized that the use of computers in building and evaluating theories of neurophysiological and psychological phenomena is a trend in no way out of the ordinary: the great majority of theories in physics and chemistry are built up and evaluated on computers, and similar habits are now increasingly permeating the biological sciences. Biologists generally accept that computer-based theories in their field, far from implying any disrespect to the special characteristics of living matter, may have quite as much value as in physics and chemistry for stimulating understanding and suggesting new kinds of experiment - provided only that the theoretical work takes proper account of available observational data.

Category C is especially concerned with theories to interpret neurobiological data on specific areas of the CNS, using computer-based models of neural nets to test out particular hypotheses on (say) the functioning of the cerebellar cortex. Other theories, of (say) parts of the visual cortex, may seek to relate both to neurobiological and to psychophysical data. Generally speaking, mathematically educated persons may be most effective in this field after prolonged study of CNS anatomy and physiology. Conversely, experimental psychologists and neurophysiologists may in several cases become expert in the construction of computer models from which new theoretical concepts may develop.

Other important aims in category C include the development of computer models related to observations of a strictly psychological nature, such as data on visual pattern recognition and scene analysis, on visual and auditory memory, on general aspects of associative recall. A further series of aims refer to specifically human types of CNS activity: thus, psycholinguistic studies are concerned with theories of the psychological processes concerned in the use of language, while other studies probe similarly the processes involved in classification and inductive generalisation. These are areas where the computer-based models of neural activity are inevitably remote from the hard facts of neurobiological observation, but where contact with the data of experimental psychology is of crucial importance.

Some workers in this field identify the essential long-term aim as [understanding the human intellect](#), but they mean this only in the sense that the aim of cosmology is [understanding the past, present and future of the universe](#). There is no implication that such generality is apparent in anyone group of research lines, merely that a general direction of desired improvement of knowledge is common to many such groups.

One more group of category C researches is concerned with how the human intellect acquires knowledge and skills, and this is related to educational psychology. For example, behavioural data on the order of acquisition of different abstract concepts in childhood may be studied in relation to models for the structuring of such concepts within the CNS.

To sum up, category C is concerned with basic research on Computer-based studies of CNS function, including the function of particular areas like the cerebellar cortex or parts of the visual cortex, and also special functions like visual pattern recognition and scene analysis, visual and auditory memory, associative recall, psycho-linguistics, classification, inductive generalisation and learning. This is work essentially within the life sciences and involving the pursuit, for its own sake, of knowledge which must appear to us as introspective living beings particularly desirable of attainment.

Evidently, there is a vast difference of approach between the practical, technological aims of category A (Advanced Automation of human activities) and the fundamental, biological aims of category C (Computer-based CNS studies). The aims are in each case perfectly clear, and perfectly distinct. The affinities in each case are much stronger with neighbouring fields (category A with general computer science and control engineering; category C with general neurobiology and psychology) than with each other. The appearance of a few common terms among the interests within the two categories (for example: pattern recognition, linguistics, inductive generalisation and learning) does admittedly indicate a degree of overlap, but may exaggerate its extent, as the problems of simulating these functions to achieve practical aims are not necessarily at all like the problems of studying how the CNS achieves them. If categories A and C were the whole body of research with which we had to deal we would recognise a minor extent of overlap of interest but regard the two areas of work as quite sufficiently distinct to warrant completely separate treatment in respect of research support, departmental organisation, etc.

Category B

Thus, the whole case for the existence of a continuous, coherent field of Artificial Intelligence research (AI) depends critically on whether between categories A and C there exists a significant category of research that may be described as a [Bridge](#) category, B, as well as on the strength of the case for any researches in that category. The existence of research work in this category is hardly in dispute: such work, as stated earlier, has been voluminous for many years, but there are much greater difficulties in any attempt at clear identification of good reasons for putting resources into those researches. The activities and stated aims of work in category B are described in the remainder of [section 2](#).

Here, letter B stands not only for [Bridge activity](#), but also for the basic component of that activity: Building Robots. The whole concept of Building Robots is, indeed, seen as an essential Bridge Activity justified primarily by what it can feed into the work of categories A and C, and by the links that it creates between them.

Thus, a Robot in the sense used here, and by most workers in the field, is an automatic device that mimics a certain range of human functions without seeking in any useful sphere of human activity to replace human beings. Work in category B (Building Robots) is frequently justified because it simultaneously supports category A (Advanced Automation), in the sense that generalised information on automatic devices may emerge which can be used in practical problems of Automation, and supports category C (Computer-based CNS studies), in the sense that devices that mimic a human function may assist in studying, and in making a theory of, that function.

These are serious arguments, that will need to be considered seriously in [section 3](#) and [section 4](#). On the other hand, they are probably by no means the only reason why Building Robots is a popular activity. At the other extreme of the spectrum of reasons we have to remember the long-standing captivation of the human imagination by the very concept, as shown by its continual prominence in literature, from medieval fantasies of the Homunculus through Mary Shelley's [Frankenstein](#) to modern science fiction. To what extent may scientists consider themselves in duty bound to minister to the public's general penchant for robots by building the best they can?

Incidentally, it has sometimes been argued that part of the stimulus to laborious male activity in [creative](#) fields of work, including pure science, is the urge to compensate for lack of the female capability of giving birth to children. If this were true, then Building Robots might indeed be seen as the ideal compensation! There is one piece of evidence supporting that highly uncertain hypothesis: most robots are designed from the outset to operate in a world as like as possible to the conventional child's world as seen by a man; they play games, they do puzzles, they build towers of bricks, they recognise pictures in drawing-books ([bear on rug with ball](#)); although the rich emotional character of the child's world is totally absent. Builders of Robots can justly reply that while robots are still in their infancy they can mimic only pre-adult functions and a limited range of those at most, and that these will lead on to higher things. Nevertheless, the view to which this author has tentatively but perhaps quite wrongly come is that a relationship which may be called pseudo-maternal rather than Pygmalion-like comes into play between a Robot and its Builder.

General aspects of work in category B involve work on mimicking some special functions that are particularly highly developed in man: co-ordination of eye and hand; visual scene analysis; use of natural language; [common-sense](#) problem solving. These areas for work in category B are evidently well chosen for giving good chances of feeding valuable results into the work of categories A and C.

Various reasons including limitations of computer power have restricted the [universe of discourse](#) in which the functions just mentioned are exercised in existing robots to something like a chessboard, or a simple [table-top world](#) on which coloured blocks are moved about and stacked on one another. Several workers have argued that games such as chess and draughts are ideal spheres for development of robot potentialities because there is great scope for ingenuity but little waste of programming effort on inessential features resulting from too extensive a universe of discourse.

To sum up, category B is a Bridge Activity concerned with Building Robots for purposes which include the feeding of information into the work of categories A and C; each Robot is designed to mimic some group of human functions, including functions such as eye-hand co-ordination, scene analysis, use of natural language, problem solving, etc, within some limited universe of discourse such as we may exemplify by a game (chess, draughts, etc), a puzzle, a table top on which blocks are moved about, or a drawing-book. One's views of the fundamental coherence of the whole field of AI spanning categories A, Band C must depend on one's opinion on whether the arguments for this Bridge Activity in category B are sound enough for it to be regarded as a necessary concomitant to, and link between, the rather different and rather easily defensible activities in categories A and C.

3 Past disappointments

Most workers in AI research and in related fields confess to a pronounced feeling of disappointment in what has been achieved in the past twenty-five years. Workers entered the field around 1950, and even around 1960, with high hopes that are very far from having been realised in 1972. In no part of the field have the discoveries made so far produced the major impact that was then promised.

The disappointment felt may be analysed into two kinds: work in the categories A and C of [section 2](#) has some respectable achievements to its credit (and achievement in such categories of work with rather clear aims is clearly discernible), but to a disappointingly smaller extent than had been hoped and expected, while progress in category B has been even slower and more discouraging, tending (as explained in [section 2](#)) to sap confidence in whether the field of research called AI has any true coherence. In the meantime, claims and predictions regarding the potential results of AI research had been publicised which went

even farther than the expectations of the majority of workers in the field, whose embarrassments have been added to by the lamentable failure of such inflated predictions.

These general statements are expanded in a little more detail in the rest of [section 3](#), which has been influenced by the views of large numbers of people listed in [section 1](#) but which like the whole of this report represents in the last analysis only the personal view of the author. Before going into such detail he is inclined, as a mathematician, to single out one rather general cause for the disappointments that have been experienced: failure to recognise the implications of the [combinatorial explosion](#). This is a general obstacle to the construction of a self-organising system on a large knowledge base which results from the explosive growth of any combinatorial expression, representing numbers of possible ways of grouping elements of the knowledge base according to particular rules, as the base's size increases.

Category A

Achievements within the sphere of the Advanced Automation (category A) have to be judged in competition with what industry has been able to achieve during the same period by perfectly conventional methods of control engineering and data processing. We may remind ourselves of the toughness of this competition by two examples. The human skills required to land a large aircraft reliably and safely are complex and intricate; yet the Automatic Landing System of Smith's Aviation Ltd., which uses classical control technology, has a better than human performance and has now been certified by the Air Registration Board, which for the purpose had demanded to be convinced of a less than 1 in 107 failure rate. Another British firm, Image Analysing Computers Ltd., has had a considerable commercial success using conventional programming methods to analyse images (eg microscope slides) as scanned by a television raster and to give numerical data (eg on metallographic grain shapes and sizes, or on cell characteristics in blood samples) without human intervention; automatic cervical-smear analysis now seems achievable by these means.

Workers in category A, while recognizing the effectiveness of such conventional control-engineering and data-processing methods applied to particular specialised tasks, have tended to emphasize the likelihood of Advanced Automation techniques of far more general applicability emerging from their work. The concept of automatic devices or methods with general capabilities is certainly a most attractive one. It is therefore particularly disappointing that the experience of the last twenty-five years has increasingly forced workers in category A to conclude that Advanced Automation techniques are successful not when they are developed with a high degree of generality of application, but only when a large quantity of detailed knowledge about the problem domain is utilised in the program design.

While this conclusion, which is rapidly gaining acceptance, has been undermining one of the clearest overall justifications for work in category A, performance of Advanced Automation systems developed at great expense in problem domains of particular economic importance has generated a still stronger sense of disappointment. Work in the pattern-recognition field has not yet proved competitive with conventional methods: even the recognition of printed and typewritten characters posed a quite surprising degree of difficulty, while the recognition of handwritten characters appears completely out of reach. Speech recognition has been successful only within the confines of a very limited vocabulary, and large expenditure on schemes to produce machine recognition of ordinary speech has been wholly wasted. Learning techniques, by which a machine's performance at recognising words might improve on receiving identified words from more and more individual speakers, appear feasible only for an exceedingly small vocabulary (such is the power of the combinatorial explosion) like the decimal digits!

The most notorious disappointments, however, have appeared in the area of machine translation, where enormous sums have been spent with very little useful result, as a careful

review by the US National Academy of Sciences concluded in 1966; a conclusion not shaken by any subsequent developments. Attempts based on classical grammar and syntax and on the transformational grammar of contemporary general linguistics have been equally unsuccessful in producing acceptable programs. Suggestions from recent research (see below), that analysis and use of natural language by computer succeed only when a very detailed knowledge of the universe of discourse is stored within the machine, augur badly for the future availability of machine-translation programs versatile enough to be commercially valuable.

Mathematical theorem-proving is another area of work in category A that has had its disappointments. Of course, conventional programming is used by many pure mathematicians with great success to generate examples suggesting, or counter-examples disproving, theorems; while conventional proofs that leave a finite residuum of cases unaccounted for may often be completed by a computational survey of those cases. The mathematician is then using the computer as a fast, reliable and [biddable](#) number-cruncher, the role in which computers generally have been most successful.

In the nineteen-fifties and early nineteen-sixties, however, a great deal of optimism was generated from the concept of realising on a computer the algorithms for theorem proving suggested by the decidability propositions of mathematical logic, starting with the [completeness theorem](#) of the first-order predicate calculus. Those most involved now emphasize that this is particularly an area where hopes have been disappointed through the power of the combinatorial explosion in rapidly cancelling out any advantages from increase in computer power. The modern trend is to [heuristic](#) methods, which also are the only methods that have been found effective in the general areas of problem solving and graph traversing.

It is important to understand the meaning attached to this adjective [heuristic](#) which increasingly permeates the Artificial Intelligence literature: it means that the program stores and utilises a large amount of knowledge derived from human experience in solving the type of problem concerned. Thus, it depends critically on data derived through the use of human intelligence, so that the widespread view that only heuristic methods are effective is a serious setback to more extreme versions of the AI philosophy. For example, a heuristic graph-traversing program requires stored values of a human estimate of the [nearness](#) of each node to the desired goal.

Some interesting concepts from mathematical logic that have been influential on work in category A are those relating to two or more statements including variable elements: namely, their [resolution](#) or greatest common instantiation, and its [dual](#) their least common generalisation which represents a sort of [inductive inference](#) from them. Algorithms exist for obtaining these but limits on their practical use again result from the combinatorial explosion.

There are, in addition, difficulties in using the techniques of mathematical logic in heuristic programs based on stored knowledge, particularly because the type of stored knowledge favoured by logicians, namely a set of axioms, is inconvenient for access by practical programs!

An excellent example of successful work in category A that has resulted from storage of as much detailed information as possible about the problem domain is the [heuristic dendral](#) program for inference of chemical structure from mass-spectroscopy data. Its output is a list of possible molecular [graphs](#) (ie structures) in order of decreasing plausibility that are consistent with the mass spectrum and the empirical formula and in some cases data of certain additional types. It has been the extremely careful study of extensive detailed information affecting the relationship of chemical structures to mass spectra that has brought about the relatively good success in this field.

In just the same way, quite good performance has been achieved in complicated areas of data storage and retrieval where the problems were confined to data of very precisely defined and analysed types. By contrast, generalised information-retrieval systems have been somewhat disappointing, especially when applied to research information involving relatively advanced ideas.

To balance seriously limited successes in achieving the longer-term objectives of work in category A, one must recognize a great deal of **spin-off** from such work, and from associated work in category B, into the **software industry** and into programming technique generally. Certain high-level programming languages developed for this work have proved invaluable in a wide range of programming activity. The list-processing languages have many advantages over conventional programming languages; for example they eliminate the labour of preliminary estimation and organisation of store space. Languages specially suitable for problem solving and for linguistic analysis have also been derived. Their advantages include **automatic back-tracking** by which if a particular subroutine fails all activity is **unwound** back to a specific point and then an alternative subroutine is tried. There is a very widespread appreciation of the many merits of this group of programming languages. It must be admitted, on the other hand, that excellent work on developing high-level programming languages has been done also in regular computing laboratories and in research groups devoted to general computational theory.

Category C

The history of work in category C (Computer-based CNS Studies) has been somewhat similar to that in category A: in spite of a respectable volume of achievement resulting from such studies, most workers who entered the field around ten years ago confess that they then felt a degree of **naive** optimism which they now recognise as having been misplaced. It is, once more, the most generalised types of studies whose end-products have proved most disappointing.

There is a consensus of view that benefits from this work in category C have flowed primarily to the science of psychology: in fact, a new range of attitudes to psychological problems has been generated. Computer models, although yielding no sudden breakthrough, have helped distinguish between theories of psychological phenomena which are possible candidates for consideration and theories that simply cannot be made to work.

As might be expected, some of the best work is by actual experimental psychologists with a good knowledge of a complex mass of data who have acquired the skills needed to build computer models for interpreting it: such work developed, for example, the concept of the visual buffer store. One school of thought emphasizes the value of intimate relation of computer models to detailed CNS data so exclusively as to propose denial of computer capacity to more theoretical groups until the demand for computer capacity from such experimental psychologists and from neurophysiologists is fully met.

Another school of thought sees a real place for the more speculative theorists, however, and points out the potential value of current studies of the types of neural networks that might be effective for functions such as associative recall, classification by attributes and inductive generalisation. It is easy to believe that, as in physics and chemistry, the more speculative theorists do have a real role to play in generating ideas. On the other hand, some of the most significant work in these neural-net theories has been done in close association with local neurobiological data. Furthermore, some of the most remarkable neurobiological discoveries, including many on the structure of the visual cortex, have not required any computer-based modelling at all! A properly balanced view of work in category C may perhaps be that the besetting applied-mathematics sin of taking insufficient trouble to master the experimental facts needs to be carefully guarded against but that, if it is, the work

produced can significantly help in the long process of moving towards better understanding of CNS function.

Psycho-linguistics is an area of psychology where this may particularly be the case. The algorithmic approach to the subject apparent already in transformational grammar and its syntactical theories of how sentences are generated is now being extended to involve algorithms taking into account more semantic information; that is, more knowledge about the universe of discourse. This type of algorithm looks much more promising as a model of how the CNS processes language.

The area nearest to an applied science which we listed in [section 2](#) as coming within category C was educational psychology. There has recently been speculation on whether the time may be ripe for research aimed at direct application of AI research to educational method through the development of advanced forms of Computer Aided Instruction (CAI).

There is a well established, [classical](#) approach to CAI that gives quite good results in educational areas that may be described politely as [drill and practice](#), or less politely as [cramming](#)! The [teaching machine](#) is programmed to print out factual information interspersed with multiple-choice questions and to go into various alternative loops, in which it prints out encouraging or corrective comments (with additional questions in the latter case), according as the right answer or one of the wrong answers is chosen.

There are those who hope to go beyond this type of CAI to a type that might be suitable for a wider range of material than mere cramming of facts and might respond more sensitively to the abilities and difficulties of the pupil. It might depend not only on pre-stored material but also on a programmed capability to generate new material, using natural language, from a stored [knowledge base](#).

These applied-research dreams are in the present author's view singularly untimely and unpromising. Taking into account the very large computer capacity and programming skill needed at present to achieve computer use of natural language on even a very small knowledge base, and adding all the difficulties of structuring and accessing a larger knowledge base and monitoring pupil performance, one can only conclude that the nineteen-seventies are not the right decade in which to begin researches aimed at applying such techniques to the teaching of any body of knowledge big enough to be of practical interest. To avoid misunderstanding however, one should make clear that basic research on developmental psychology by the methods of category C would not on this argument be excluded.

Category B

The balance between numerous disappointments and certain solid achievements from work in categories A and C is, perhaps, typical of scientific research as a whole. It indicates only that these areas of research are not in one of those conditions of exceptional fruitfulness when everything seems to be going right. By contrast, the sense of discouragement about the intended Bridge Activity of category B, centred upon Building Robots, seems altogether more widespread and profound, and this raises doubts about whether the whole concept of AI as an integrated field of research is a valid one.

Quite possibly the sense of discouragement is greater in category B because still greater expectations have been sensed and voiced in this category than in the others. Some workers in the field freely admit that originally they had [very naive](#) ideas about the potentialities of intelligent robots, but claim to recognise now what sort of research is realistic. In these circumstances it might be thought appropriate to judge the field by what has actually been achieved than by comparison with early expectations. On the other hand, some such comparison is probably justified by the fact that in some quarters wild predictions regarding the future of robot development are still being made.

When able and respected scientists write in letters to the present author that AI, the major goal of computing science, represents [another step in the general process of evolution](#); that possibilities in the nineteen-eighties include an all-purpose intelligence on a human-scale knowledge base; that awe-inspiring possibilities suggest themselves based on machine intelligence exceeding human intelligence by the year 2000; when such predictions are made in 1972 one may be wise to compare the predictions of the past against performance as well as considering prospects for the realisation of today's predictions in the future.

It certainly seems that early enthusiasm for programming and building a robot that would mimic human ability in a combination of eye-hand co-ordination and [common-sense](#) problem solving has ended up gravely disappointed. The large amount of computer time needed to distinguish between everyday objects of markedly different shapes against a far from noisy background has been most discouraging; the engineering complications required to achieve eye-hand co-ordination (not of human standard but similar to what an octopus can learn) have been repellingly formidable. Reports from the world's different centres for this work are all disenchanting.

Some able research workers, who from their beginning in the field regarded Building Robots as a precarious or even [crazy](#) enterprise but nevertheless were attracted to participate in such a long-shot or even [shot in the dark](#) activity, have felt themselves driven now to recognise that the difficulty of achieving good hand-eye co-ordination in quite simple problem situations has proved unexpectedly great, and seems to hold out negligible hope of approaching human levels of achievement. In these circumstances, many good computational theorists are emphasizing that productive research on [robot reasoning](#) (or, essentially, commonsense problem solving) does not necessarily need the physical presence of an eye-hand machine. This line of argument then branches in two directions, one leading to work properly in category A (directed to automating the solution of such problems as may arise in practical fields of application), and the other (which is our concern here) leading to programs for problem solving in an abstract [play](#) situation: for example, in an abstract table-top world with data fed in not as television images but as statements about the positions of blocks on the table-top; or in a similarly defined chessboard or puzzle situation.

The [Category B](#) research work on problem solving in these abstract play situations has produced many ingenious and interesting programs. A fair description of the success of these programs seems to be that they are effective when and only when the programming has taken into account a really substantial quantity of human knowledge about the particular problem domain. Just as in category A, the pure mathematical logic methods suffer defeat at the hands of the combinatorial explosion, and have to be replaced by [heuristic](#) methods. Some very interesting researches have been carried out to develop [general](#) problem-solving programs, and such work can be of research interest to psychologists, but the performance of these programs on actual problems has always been disappointing. Students of all this work have generally concluded that it is unrealistic to expect highly generalised systems that can handle a large knowledge base effectively in a [learning](#) or [self-organising](#) mode to be developed in the 20th century.

Those wishing to decide as between this view and the quite opposite views of the [awe-inspiring](#) future mentioned earlier can quite helpfully study the state of the art on chess-playing programs. This is partly because chess is a complicated enough game so that in a contest between a computer and a human player the computer's advantages of being able to calculate reliably at a speed several orders of magnitude faster need by no means be decisive (the number of possible positions being incomparably greater) and so there is real interest in whether or not they are outweighed by the human player's pattern-recognition ability, flexibility of approach, learning capacity and emotional drive to win. Another good reason for investigating chess-playing programs is that the long-term interest of the big

international computer manufacturers in bringing about some spectacular achievement of [machine intelligence](#) against such a well developed human intelligence as an able chess player, in order to assist in selling more generally their products' potentiality for superseding human intellectual activity, has been an incentive to the devotion of quite considerable resources to producing an effective program.

It is interesting to consider the results of all this work some twenty-five years after the researches aimed at chess-playing programs began: unfortunately these results are discouraging. The best programs play chess of only [experienced amateur](#) standard characteristic of county club players in England. Chess masters beat them easily.

More important, progress on constructing chess-playing programs has been made solely by heuristic methods. The programs seek to maximise in what may be called the foreseeable short term a complicated [evaluation function](#); this function, constructed entirely from human knowledge and skill, represents an evaluation of a position, depending on large numbers of different measurable features of it with different weights attached to them. What relatively modest success the programs have achieved is a measure primarily of human skill and experience in the concoction of this evaluation function. The computer's contribution is primarily rapidity in looking a few moves ahead and finding a line that produces a position change good on the basis of that evaluation. The [intelligence](#) contribution is human; what the computer offers is its speed, reliability and biddability. By contrast, [learning](#) programs are not considered applicable to computer chess at present.

To sum up, this evidence and all the rest studied by the present author on AI work within category B during the past twenty-five years is to some extent encouraging about programs written to perform in highly specialised problem domains, when the programming takes very full account of the results of human experience and human intelligence within the relevant domain, but is wholly discouraging about general-purpose programs seeking to mimic the problem-solving aspects of human CNS activity over a rather wide field. Such a general-purpose program, the coveted long-term goal of AI activity, seems as remote as ever.

In thus regretfully noting the remoteness of this goal we must not, however, be tempted into overvaluing it because of its inaccessibility. We must remember, rather, that the [intelligent problem solving](#) and [eye-hand co-ordination](#) and [scene analysis](#) capabilities that are much studied in category B represent only a small part of the features of the human CNS that give the human race its uniqueness. It is a truism that human beings who are very strong intellectually but weak in emotional drives and emotional relationships are singularly ineffective in the world at large. Valuable results flow from the integration of intellectual ability with the capacity to feel and to relate to other people; until this integration happens problem solving is no good because there is no way of seeing which are the right problems. These remarks have been included to make clear that the over-optimistic category-B-centred view of AI not only fails to take the first fence but ignores the rest of the steeplechase altogether. It will suffice, however, to judge the work on its own rules and its own aims in order to conclude that the attempt to construct a true Bridge between categories A and C is not succeeding.

Postscript

It is only fair to add at the end of this section on [Past Disappointments](#) that some workers in the field would have agreed with the view just expressed until the appearance less than two years ago of an exceptionally good PhD thesis ([Understanding Natural Language](#) by Terry Winograd, MIT. Published in UK by Edinburgh University Press, 1972) on a computer program for use of natural language, since when they have felt resurgence of optimism about the coherence and viability of the concept of integrated AI researches. It is important to analyse how this reaction has come about and how far such resurgence is justified.

The thesis is exceptional in more than one way. The style in which most papers on AI research are written is depressingly turgid or jargon-dominated and almost makes the authors appear antagonistic to the special human gift for relating to, and communicating with, other people in an imaginative way (as if such authors appreciated only those human capabilities which they seek to mimic in robots!) By contrast, the thesis is a pleasure to read, the author's substantial research achievement and attractive personality being communicated infectiously by his style of writing. His gift for language and communication has without doubt contributed to making his researches widely known all over the world, as well as having contributed to the success of the actual linguistic analysis underlying the development of his program.

This analysis is the strongest of those mentioned above under category C as having developed far beyond the transformational grammar approach of general linguistics into new methods for machine interpretation of natural-language sentences within a limited universe of discourse, that make very substantial use of stored knowledge concerning that universe. Full use of such knowledge is regarded by the author of the thesis as an essential ingredient of the success of his approach, to which however the penetration and originality of the analytical methods he introduced has also made a vital contribution.

Specifically, the universe of discourse is an [abstract table-top world](#) and in the lengthy, and now rather famous, [conversation](#) between the author and his program the program accepts, and is deemed to have carried out, certain commands to perform well defined block-stacking operations, while it queries commands that are impossible or ambiguous. The program deals similarly in answer to questions put to it regarding the present and past states of the table-top world. In constructing the program, two of the high-level programming languages referred to under category A above were used: one to program the events in the abstract table-top world and one to perform the linguistic analysis. The thesis well illustrates the value of these high-level languages.

The contribution of brilliant presentation and deep originality has made this thesis deservedly influential on workers in most areas of AI research. The biggest and clearest influence is on psycho-linguistics itself (work in category C), where the studies have been in large part responsible for a movement towards viewing the processes by which the human CNS responds to and uses language as semantics-controlled (or knowledge-controlled) even in their fine structure.

Many workers in category B (the Bridge Activity of Building Robots) have at the same time felt encouraged by this thesis: its program, after all, can properly be described as a Robot [with whom the author converses](#), and Building this Robot has succeeded in its aim to an extent undreamt-of in the unrewarding world of eye-hand machines. The program seems furthermore to open up more general possibilities of conversing with Robots by means of natural language. There is even a suggestion of an ultimate link through to work in category A if these new studies could revive prospects for achievement of machine translation.

To such somewhat over-generalised euphoria it is necessary to respond, however, with certain cautionary reservations beyond the banal comment that one swallow does not make a summer. Outside the psycho-linguistic area where the thesis has truly helped to establish a new direction of research, suggestions for possible developments in other areas that can properly be inferred from the studies are rather discouraging. Thus, the studies show how the complex problems involved in computer use of natural language are rendered far more complex by the need to interact in detail with systems for structuring and accessing the necessary knowledge base. For an extensive universe of discourse this could put such developments out of practical reach.

In practice, a large computer together with very sophisticated programming using subtle new programming-language developments was found just sufficient to make slow conversation

possible on the very limited material represented by the abstract table-top world; material restricted enough, for example, to allow resolution of ambiguities in natural-language sentences by classical theorem-proving techniques. Extension of the methods used to a much wider universe of discourse would be opposed violently by the combinatorial explosion.

Accordingly, the present author's view of the definite (though not overwhelming) promise of work in categories A and C, and of the general failure of work in category B to establish effectively the unity of AI research as a whole, remains unmodified by careful study of one particular piece of work of a very remarkable character. This postscript to the section on [Past Disappointments](#), explaining this, has been included because emotional response to such work is very natural and desirable but needs to be integrated properly with an intellectual appraisal of where its significance primarily lies. The thesis is, of course, a triumph of human intelligence, and human intelligence can respond to it most correctly by recognising its main contribution as being to aspects of how the use of language by the human race has developed and of what processes within the human CNS that use may involve.

4 Future possibilities

However controversial may be an analysis of the past, a forward look towards the different possibilities that the future may hold in some area must be more controversial still, especially when made in a report to a body whose decisions can have a very substantial influence over a certain part of that area's future (specifically, the British part). Controversy has its unattractive as well as its attractive features, but it cannot be avoided in a period when the abilities of scientists jointly to arrive at wise decisions on research policy is publicly regarded as being on trial.

Research on AI in some other countries may be funded by military agencies (ARPA in USA) or by other mission-orientated public bodies. With this type of funding it is common for scientists to [close their ranks](#) and avoid public disagreement among themselves, in the hope that the total funds available for science may thus be enhanced to an extent that may outweigh any harmful results of a distribution of those funds determined on the basis of insufficient scientific discussion. Such optimism would be unjustified in a poorer country such as Britain, while the alternative approach here advocated accords with the desire to [keep our AI research civilian](#) expressed to the author by various British workers in the field. This suggests that decisions within the UK should be taken only after carefully contrasting and comparing different informed views of the research field's future available to SRC. Thus, due weight should be given to the principle [Heterarchy not Hierarchy](#) (an AI maxim of considerable soundness concerned with file structures).

To the required debate this report's contribution consists not of any detailed costed recommendations, but of certain general considerations based on the analysis of the past given in [section 3](#) as well as some thoughts about the present and future now to be presented. After omission in this published version of all specific comments on British research work in the field, these consist essentially of an attempt to look to the field's scientific future in the world as a whole, subject to the proviso that any speculations beyond the end of this century are regarded as too uncertain to justify mentioning them or basing on them any present research decisions whatever.

It is assumed that more precise policy formulation and detailed decisions on projects will stem from the normal working of the machinery of the Science Research Council and its Boards and Committees, influenced to some extent by such special reports as may be available, including the 1972 Policy and Programme Review of the SRC Computing Science Committee and also the 1972 report of a joint SRC/MRC panel on Neurobiology, as well as the present [personal view](#) of AI as a research field.

[The next twenty-five years](#)

This [personal view](#) which saw the past twenty-five years of AI research as having developed a [bimodal](#) distribution of achievement, with some respectable (though not as yet lofty) peaks of achievement in categories A and C but relatively speaking a valley between them in category B, looks ahead to still greater bimodality, amounting practically to fission, arising during the next 25. Specifically it foresees, whether within category A or category C, certain research areas making very substantial further progress, coupled in each case with the forging of far stronger links to the immediate field of application than to the supposed bridge activity B. Rising confidence about the work's relevance within the associated field of application may add prestige and thence strength to such an area of research, while continued failures to make substantial progress towards stated aims within category B may cause progressive loss of prestige, from which a diminution of funding will ultimately follow even where scientific claims are not always subject to full scientific scrutiny. In due course the overriding significance of the links between each research area and its field of application will rupture the always fragile unity of the general concept of AI research.

For example, in the technological applications within category A the work will become increasingly specialised, in accordance with the common experience that AI researches are successful to a degree closely correlated with the degree of use made of detailed knowledge of the problem domain. Techniques for Advanced Automation can now be expected to move forward fastest where research workers concentrate upon practical problems, acquiring for the purpose good detailed knowledge of the technological and economic contexts of the problems chosen. Benefit to both sides will flow from closer integration with control engineers, who have deep experience not only of the practical but also of the financial and sociological questions arising in automation, and can conversely learn much that is of value to them from experts in advanced computational theory. That theory itself (including the development of new programming languages) can in the meantime forge ahead through a combination of the spin-off from Advanced Automation developments and of the activities (serving far more than AI research) of general computing-science laboratories.

Recently the Japanese government announced a £40M research programme oriented very closely along these lines, aimed at the automation of factory assembly processes, which among the areas within category A mentioned in [section 2](#) may well be one of the most promising. Another clue to how Advanced Automation developments may proceed comes from recent experience in the sister field of Computer Aided Design (CAD): generalised researches dominated early CAD work, but later on several specific industries developed forms of CAD very specialised towards their own problems. This example is mentioned mainly as an analogy to how Advanced Automation may come to experience a similar degree of fragmentation; however, we should also remember from [section 2](#) that CAD is itself an area where advanced computational theory may be found to have a role, especially in the integration of subsystem designs. Close relationships between AI theorists and CAD workers (such as here and there have already come about) are one prerequisite for those developments.

A similar outward-looking trend is expected in the mathematical and scientific applications of researches within category A; in mathematics, for example, from utilisation of far more detailed observation of [how mathematicians actually prove theorems!](#) The structuring and utilisation of scientific data bases is another area where good results depend on detailed study of the data's special characteristics. The one part of that field with which the present author has been closely involved, as Chairman of the Steering Committee for the Experimental Cartography Unit of NERC since its inception, affords a good example of this: the structuring of geographically located data is found to demand quite specialised techniques, closely related to the cartographic character of the output. Another good example from the scientific sphere is the chemical-structure work described in [section 3](#).

It is not to be expected that, in all the areas within category A listed in [section 2](#), striking successes will be reached during the next twenty-five years. The view here proposed is rather that the chances of success in anyone area will be greatly improved through close integration of the researches with the field of application.

Substantial advances are at the same time expected within category C, where success will again be related to how closely the work is linked to the fundamental associated disciplines of psychology and neurobiology. Computer-based CNS studies can from experimental psychology gain greatly through more substantial use of the extensive data available, eg on reaction times, on pattern-recognition abilities, and on the types of errors made in different tasks. From modern neurobiology they can derive a valuable appreciation of the detailed evidence to the effect that the CNS uses [specialised hardware](#) very economically to perform significant and important tasks. Computer-based studies have a role to play in analysing how some of this hardware may function, and conversely can derive a salutary reminder that simulation of the extraordinary self-organising capabilities evolved by the human CNS may actually be unattainable through ingenious software developments.

Conversely, psychology and neurobiology will benefit to an extent closely related to how far Computer-based CNS researches behave as if they felt integrated within one or both of those fields. Psychologists and neurobiologists may especially gain increased appreciation of the value of computers for theorising about complex systems and for making sense of complex masses of data, while integration at a more fundamental level can be expected to follow. With the resulting growth in understanding of the human CNS, respect for it may, perhaps, grow to an extent that will reduce the ebullience characteristic of past predictions of AI possibilities.

In the meantime, the intended Bridge Activities within category B may well have been found increasingly disappointing as achievements from Building Robots of the more generalised types fail to reach their more grandiose aims. On the other hand, some robots designed primarily as computer models for comparison with experiments on how the human CNS performs linguistic or problem-solving tasks will become integrated with work in category C, while others aimed at practical tasks, related eg to engineering assembly, will become integrated in category A.

These processes are expected to bring about, at a slow but increasing rate, the fission of the field of AI research predicted at the beginning of this brief attempt at looking into the future. That attempt may now be concluded with the observation that such a broad-brush view, unspecific in matters of detail, is possibly all that can properly be attempted over time-spans as long as twenty-five years.

Part II Some Comments on the Lighthill report and on Artificial Intelligence

Professor N S Sutherland

Professor of Experimental Psychology, University of Sussex, August 1972

Introduction

No one is likely to quarrel with Lighthill's classification of work in AI as falling roughly into three areas; nor is his recommendation to support good work on automation and on the computer simulation of neurophysiological and psychological processes controversial. The crux of the report is the condemnation of work in the bridging category and I shall therefore concentrate on this aspect of it. It will be argued that Lighthill's definition of area B is misleading, that some of his arguments against work in this category are unfounded, that the achievements and promise of the work can be seen in a very different light from that which he presents them and that it is hard to see how work in areas A and C can flourish unless there is a central core of work in area B. In what follows, Lighthill's abbreviations for the three areas will be retained, but B will stand for [Basic research in AI](#) rather than for [Bridging research](#): the same work is denoted, but the connotation is different.

1 Definitions and aims of area B

In [Section 2](#), Lighthill defines area B as [a bridge category where aims and objectives are much harder to discern \(than in the case of categories A and C\) but which leans heavily on ideas from both A and C and conversely seeks to influence them](#). Later in this section we are told that the basic component of area B is building robots and that this activity is [justified primarily by what it can feed into the work of categories A and C](#). Area B is already being placed in a bad light since on the one hand it is said to borrow its ideas from A and C, and on the other hand its main justification is said to be what it can contribute to these areas. Four points arise.

First, area B has clearly defined objectives of its own. Its aim is to investigate the possible mechanisms that can give rise to intelligent behaviour, to characterise these mechanisms formally, and to elucidate general principles underlying intelligent behaviour. These seem to be valid scientific aims and are clearly different from those of work of types A and C. Workers in area B need not concern themselves with such problems as whether a given mechanism is the most economic method of doing something nor need they worry about whether it is a mechanism actually used by an organism. Lighthill gives no reasons why the theoretical investigation of intelligent mechanisms is not a useful scientific enterprise in its own right. Moreover, I believe that this view of the aims of work in the central area of AI is shared by the hard core of workers in the subject.

Secondly, Lighthill suggests that so far work in area B has received more benefit from ideas emanating in areas A and C than vice versa. No evidence is given in support of this view and it could indeed be argued that so far almost all the spin-off has been from the central area (B) to A and C rather than the other way round. The new concepts central to all work in AI (eg list processing, backtrack, heterarchy not hierarchy, knowledge as procedures) have developed from work in B rather than from work in areas A and C. Moreover, the family of languages known collectively as list processing languages was developed almost entirely by workers in area B and these languages have been of use to workers in areas A and C and to computing science as a whole.

Thirdly, the categorisation of all work in area B as [robotics](#) is tendentious. The word robotics is best kept for [hand-eye](#) projects and although Lighthill mentions work on scene analysis,

question and answering, and the understanding of an input in a natural language as belonging to area B, little attention is given to the achievements of such work; rather the alleged stigma attached to work on robotics is taken to characterise area B as a whole.

Fourthly, because of the vagueness in Lighthill's definition of area B, it is not clear who he thinks works in this area and who does not. Some of the most outstanding work in the field of AI as a whole has been achieved by workers such as Minsky, Guzman, Winograd, Winston, Evans and Raphael, all of whom would properly regard themselves as in the central area of AI; it is hard to think of any work in area A and C that is of such scientific importance. If SRC is to support one area of AI rather than another, the question of definition becomes very important. Moreover, the lack of definition and the lack of names may mean that Lighthill is attributing achievements in area B to area C. For example, it is said that there is a group at Edinburgh working in C but on the above definition all the work at Edinburgh would be in areas A and B.

2 Lighthill's arguments against work in B

Some of Lighthill's arguments against B seem rather tendentious. A listing and rebuttal of these arguments follows.

2.1 It may or may not be true that some male scientists' work on robotics is unconsciously motivated by a desire to give birth ([page 7](#)). What is certainly true is that a scientist's unconscious motives for undertaking the work he does have nothing whatever to do with the assessment of the scientific value of that work. The ascription of an amusing if somewhat derogatory motive for undertaking work on robotics would not matter if some attempt were made to assess the scientific reasons for undertaking robot work. There is a scientific justification for robotics, although it is not explicitly stated in Lighthill's report. Only by having a machine that takes an input directly from the real world and itself operates on the real world is it possible to make sure that important problems are not being overlooked and that reliance is not still being placed on the enormous amount of data processing done implicitly by humans before giving an input to the machine. An excellent example of this is the problem of [dirty](#) pictures: until an attempt was made to make machines process raw information from the real three-dimensional world, it was not realised just how messy most visual inputs are nor was the complexity of the information processing needed to interpret such input appreciated. (A milder form of the same problem had previously been encountered in attempts to make machines read print, but such attempts escaped many of the difficulties posed by the three-dimensional world - eg the problem of lighting.)

2.2 Lighthill's remarks on Winograd's work also seem to be misleading. The [remark](#) that [one swallow does not make a summer](#) is particularly out of place. This work was not as new as Lighthill makes out: Winograd used the knowledge and techniques developed over 10 years of previous work in area B, and brought them all very neatly to bear on one problem. LISP and MICROPLANNER, languages developed by others, are foreshadowed in earlier papers by Minsky, the grammatical system is similar to that of Woods, and the implementation of the program was dependent entirely of the existence of PLANNER (developed mainly by Hewitt). Winograd's thesis, therefore, rests on a solid basis of previous work and there are many other excellent pieces of work not mentioned by Lighthill which have not had the [succes d'estime](#) of Winograd's program partly because their authors were not so adept at putting their ideas across in a readily understandable way.

2.3 The fact that progress in making machines perform intelligent tasks has not lived up to some of the wilder forecasts is again irrelevant to assessing the actual merit of what has been achieved. It would indeed be remarkable if twenty-five years of work by a handful of scientists had succeeded in producing machines with an intelligence to match that achieved by man in the course of a million years of evolution. Nor is it true (as Lighthill suggests) that

many of the main practitioners of the art are **disenchanted** with the progress of AI: a visit to the MAC project at MIT, which is at present the most important laboratory of AI, would soon dispel this impression. One of the things that has emerged from the last twenty-five years' work is that the mechanisms underlying intelligence are richer and more complex than many people thought and this could be said to add to the scientific interest of the work rather than to detract from it.

2.4 Lighthill only produces one serious argument against work in Area B. He argues that the failure of workers in AI to produce programs exhibiting general intelligence over a wide field is due to the **combinatorial explosion**: the number of paths through a network increases in proportion to n^r , where n is the number of alternatives open at one time and r is the number of steps. Where there is a large knowledge base (many possible states of the system) it clearly becomes impossible to search through all possible paths in order to find the one that leads to the goal. This has of course been realised by workers in AI for years and it would be a dull subject indeed if intelligent behaviour could be achieved simply by making computers crunch through huge searches at random until a particular goal were achieved.

As Lighthill points out, the realisation that this sort of solution to the problem of intelligent behaviour is impossible (as well as scientifically uninteresting) has led to the development of **heuristic** methods. Such methods hopefully cut down the search space and so make it possible to achieve a goal with the minimum amount of random search, often at the expense of failing to guarantee a solution in the way that an algorithm would do. Lighthill levels two criticisms of this way of proceeding. First, he complains that the heuristics are supplied by the human programmer using his own intelligence and hence the program is not intelligent and secondly he seems to doubt whether even the use of heuristics will be sufficient to enable machines to deal with very large data bases because of the combinatorial explosion.

His first argument seems to me to be beside the point. The formalisation and use of heuristics in programs and the study of how to do this may have scientific value whether or not the heuristics are supplied by human intelligence. The investigation of how to use heuristics to produce intelligent behaviour is precisely one of the goals of work in category B. Good heuristics express a knowledge of a subject domain in such a way as to facilitate the solving of problems in that domain. They depend on a deep understanding of the problem domain and are by no means ad hoc tricks.

Lighthill is of course right in thinking that one of the problems confronting work in theoretical AI is that of discovering how to build into a program general heuristics such that the program can then itself develop specific heuristics appropriate to a particular problem domain. The successful solution to this problem depends on our own intelligence and on the amount of effort that is put into it. As in any scientific enterprise there can be no guarantee of success nor can there be any proof that it is impossible to succeed. What is certain is that work in category B has uncovered some very interesting methods of expressing knowledge especially in the areas of language understanding and scene analysis and progress on this problem seems to be accelerating rather than decelerating.

Lighthill's second argument is obscure and seems to be almost an act of faith; one simply cannot know at this stage how far (or how soon) it will be possible to build systems with very large knowledge bases using appropriate heuristics to overcome the combinatorial problem.

3 Evaluation of work in area B

As mentioned above, one of the most important contributions of work in category B has been the development of appropriate high level programming languages in which to program machines to perform intelligent tasks. If the subject is to progress, such languages are not merely a matter of convenience, they are a matter of necessity. Because of the way the human brain is organised, it can only keep track of about seven concepts at anyone time

and it is therefore virtually impossible to write programs to undertake highly complex operations using machine code. By developing higher level concepts that summarize whole blocks of low level operations, workers can think in terms of such concepts without having to worry about the details of how they are implemented.

Moreover, many general ideas that can be used in tackling many different substantive problems in AI are embodied in the languages that have been developed, eg list processing, backtrack, knowledge based procedures, structural descriptions. Considerable progress has been made in understanding what is involved in picture processing, the interpretation of natural language and the sorts of representation necessary to characterise meaning. Almost all the progress stems from work in the USA and most of it comes from the MAC project at MIT, Stanford University and the Stanford Research Institute.

One recent insight derived from Basic research on AI is that interpreting the meaning of any complex input, it is impossible to use a rigid step by step procedure. For example, in deriving the meaning of a sentence, it is not possible first to parse it grammatically and then to proceed to a semantic interpretation. A whole range of different strategies based on both a knowledge of grammar and a knowledge of possible semantic structures have to be applied at each step of the analysis, and where one procedure fails others have to be called.

Again the processing of a scene involving objects with plane surfaces cannot be achieved by successively mapping picture points onto lines, then discovering closed regions bounded by connected line segments, and finally mapping such regions on to a 3-D representation that interprets the lines as edges and the regions as surfaces and that specifies the relations between the surfaces (concave, convex, etc). Because the input picture (if taken from the real world) is always noisy, some lines that do not represent edges will be recovered in the line domain, others that do represent edges will not be recovered. It is therefore necessary to bring a knowledge of what three-dimensional structures are possible and plausible to bear at each level. For example, if the most plausible 3-D structure has an edge for which no corresponding line has been found, it may be necessary to call a procedure operating at the level of picture points to discover whether there is local evidence of such a line that had originally been missed.

In short, the machine must have a [knowledge](#) of what are the possible structures in different domains (picture points, lines, regions, surfaces) and must be able to bring this knowledge to bear at all stages in the interpretation. Many other varieties of knowledge have to be deployed in interpreting an input picture. For example, some scene analysis programs are provided with a knowledge of what it means for one body to support another and this knowledge is used to interpret the pictorial input as bodies and to arrive at the spatial relationships obtaining between the bodies. It is also necessary to use a knowledge of the effects of lighting in scene interpretation.

The advances made in the subject can perhaps most readily be appreciated by comparing this type of approach with the very crude approaches being made to pattern recognition twenty years ago, when it was thought that such recognition could be achieved either by a blind process of template matching or by the simple partitioning of an n-dimensional space in which the dimensions represent pictorial attributes.

Although the examples given come from work on scene analysis, the same approach is being used in other problems such as speech recognition and the interpretation of natural language. It is now accepted that to be usefully deployed knowledge must be incorporated not in the form of a static data base but in the form of procedures that can be brought to bear on the processing of input data at the appropriate points. The methods used for embodying knowledge in procedures and for integrating the different procedures so that they can be called at the right time have considerable generality. These methods are embodied in advanced programming languages such as PLANNER.

Although it remains true that no solution is in sight to the problem of how to make a program create its own heuristics, basic work in AI is at a very exciting stage at the moment and is advancing rapidly. Considerable progress has indeed been made in the two years that have elapsed since Winograd's thesis. More flexible languages than MICROPLANNER (eg BBN-LISP and CONNIVER) are being implemented. MICROPLANNER made it possible at any stage in the processing of input information to call a variety of procedures.

Instead of being called by name such procedures were called by a process of [pattern matching](#). The existing pattern of data and the immediate objectives were matched to a set of stored patterns representing the abilities of knowledge based procedures, and instead of calling procedures by name, it was possible to call the procedure that gave the best match to the existing pattern of data. If the procedure failed to help in the meaningful interpretation of the data, the data pattern could be re-entered and a different procedure tried (backtrack). However, MICROPLANNER makes it difficult to take account of the reason why a procedure fails. CONNIVER makes it easy to evaluate the reasons for the failure of a given procedure in interpreting a given pattern of data and to use this information to make an [intelligent](#) choice of the next procedure to be tried.

It is virtually impossible in a short space to give an adequate appreciation of progress in the central field of AI work, and this difficulty is compounded by the fact that I am not myself an expert in it. I believe that work in area B has led to important insights into the nature of the mechanisms that can mediate intelligent behaviour, and that although individual heuristics have limited generality, the methods by which these heuristics are put together and brought to bear on a common problem are not task specific and are of considerable general importance.

Some workers in AI see their aim as the development of formal systems in which to characterise and deploy knowledge. It is worth noting that such work has important implications for philosophy (particularly for epistemology) and for linguistics, as well as for work in the other two areas of AI to which we now turn.

4 Relations between areas B and A

Work in area A must ultimately be judged by its cost effectiveness. However, the best economic solution to a problem in automation does not necessarily involve setting up an [intelligent](#) system. For example, if we want to land a plane automatically, we do not need to construct a device with the intelligence of a pilot, all we need is a moderately complicated but totally unintelligent system to guide the plane down a beam. Where the aim of AI work is automation, the scientist has an obligation to produce the most economic solution which may be intellectually rather dull. He is not directly concerned with the fundamental problem that activates workers in area B - ie the attempt to understand the principles that make intelligent behaviour possible. Moreover, because of the pressure to produce results of practical usefulness, he normally does not have the time to think generally about the problems nor to produce the general purpose types of language that ultimately benefit all three divisions of the subject.

In contrast to this, the worker in area B is free to pick any problem that he thinks will constitute a useful test bed for the development of new ideas and approaches to the subject, and need not concern himself with short cuts that may be of great practical usefulness but do not lead to any new insights about the nature of intelligent behaviour.

It is perhaps for these reasons that despite the achievements of work in area A, general advances in our way of thinking about the problem of intelligence and almost all the important new concepts have come from work in area B not from work in area A. Moreover, as Lighthill rightly points out, it is essential for the worker in area A to have an extensive knowledge of the problem domain. Because of the specialist knowledge the problem domain

requires examples drawn from area A are rarely suitable for the purpose of teaching and developing the techniques of AI. For these reasons, I believe it would be a disaster to support work in area A to the exclusion of work in area B.

5 Relations between areas B and C

As mentioned above, Lighthill's report does not make the demarcation between areas B and C clear. Anyone working in area C must take into account what is known of the nervous system at a physiological and anatomical level and must also be concerned with what is known about the actual behaviour of organisms. Such behaviour is often highly irrational and unintelligent and unless workers are concerned to reproduce in their programs the limitations and errors made by organisms, they cannot truly be said to be working in area C.

For example, it is well established that only about seven chunks of information can be held in short term memory: the chunks can be of any complexity provided that representational structures for a given chunk have already been built up in the brain. Workers in area B are not concerned with simulating this type of limitation on the power of the human brain: they would regard it as an adventitious limitation that has little to do with the theoretical mechanisms underlying intelligence though it has a great deal to do with how those mechanisms are actually instantiated in the human brain.

Like workers in area A, workers in C are constrained by the nature of their problem domain: they are limited to considering those intelligent mechanisms that actually exist in known organisms and they do not investigate the general nature of intelligent behaviour from a theoretical standpoint. It is not clear what are the achievements of workers in area C to which Lighthill refers: the only example given is the development of the concept of [visual buffer store](#) but this concept was borrowed from computer science and the proof of the existence of such a store in man came from experiments and did not involve any computer programming. Concepts developed in area B have heavily influenced experimental and theoretical work in experimental psychology; for example, [structural descriptions](#) are made much use of in work on pattern recognition, and there has been a spate of work on semantic networks in man influenced by the ideas of Quillian and Raphael who work in area B. It is also true that there has been considerable feedback from work in experimental psychology into area B: for example many of the detailed heuristics used in scene analysis including the interpretation of shadows and the use of support and of symmetry were foreshadowed in the experimental psychology literature though it remained for workers in area B to instantiate these ideas in working programs.

Work in area B has also benefitted work in experimental psychology by showing that it is possible to put forward precise models of intelligent behaviour and hence raising the standard of precision to be aimed at. Moreover, the existence of well formulated models of intelligent behaviour must throw some light on human intelligence since it would be surprising if some of the principles and methods used in AI were not also used by the human brain; even if methods are devised that are not used by the human brain, they are likely to throw light on its workings if only by contrast in the same way as the existence of the propositional calculus has sharpened our understanding (by contrast) of the ways in which such words as [and](#) and [not](#) are used in everyday language.

In summary then, many of the discoveries made in area B have been made quite independently of work in C and there is no evidence that they would have been made if only work in C were supported. Like mathematics, AI tends to be a young man's subject and it is therefore important to train workers in it when they are young: it would be difficult to give an adequate training both in the techniques of AI and in those of psychology and neurobiology and this means that some workers should receive a training primarily in AI which once again means studying the work done in area B. One would hope that some of the workers so

trained would subsequently move into area C and contribute directly to our understanding of human and animal behaviour, but unless the central area of endeavour in AI receives support, it is hard to see where such workers will come from.

Incidentally, much of the modelling in experimental psychology undertaken over the last 20 years has been done by workers trained in conventional mathematics. The results have for the most part been singularly useless since mathematical formalisms fail to capture the significant features of intelligent behaviour. (Two exceptions should be made - some applications of control theory and of signal detection theory have been useful). One of the aims of workers in area B is precisely to establish a good formalism for dealing with the problem of intelligent behaviour. In my opinion, work in area B is likely to have an increasing impact on thinking in experimental psychology and it is providing a new and better way in which to formulate theories. Good work in C, therefore, definitely merits support.

6 The future of AI in Britain

It has been argued above that Lighthill's area B so far from being a bridging area is really the central area of progress in AI that work in this area is worth supporting in its own right and that if it is not supported areas A and C will suffer, both through a dearth of the sort of new concepts produced by workers in area B and also through a lack of trained workers in AI, since area B appears to be the most appropriate training ground for workers in all three divisions of the subject. Any judgement of the worth of a scientific field, particularly one as new and underdeveloped as AI, must of course be highly subjective, but it seems worth presenting a different point of view from that which appears in the Lighthill report. However, even if it is accepted that area B is worth support, it is not easy to decide how such support should be given in Britain.

The problem is that there is little first class work on basic AI in progress in Britain at the moment; much of the work is about seven years behind that being undertaken in the three main laboratories in the States and shows little sign of catching up. It is almost impossible to get an adequate training in the subject in Britain. The one postgraduate course (at Edinburgh) places too much emphasis on conventional mathematics and logic and does not capture the essence of the subject as understood in the US. There is therefore a real chicken and egg problem. If area B is important and exciting in its own right, then something should be done to remedy this; moreover, the absence of good work in area B means that areas A and C which Lighthill agrees merit support are starved of good workers. The shortage of first class workers in pure AI means that it would not be profitable to spend a large sum of money (eg £1,000,000) immediately. The following steps could however be taken:

6.1 Good young workers should be encouraged to study in the main US laboratories (particularly Minsky's laboratory at MIT) and to return here.

6.2 It is worth considering the possibility of attracting some talent from the States to help in getting the subject off the ground here.

6.3 Steps 1 and 2 would be greatly facilitated if there were in existence in Britain machines accessible to AI workers that could accept AI software written in the United States. It is impossible to attract good American workers for short or lengthy periods unless they can run their existing programs and use the languages they have developed; similarly, it would be very discouraging for a British student to return here after working in America if he was not able to build on the work he had undertaken there for lack of a suitable machine. Moreover, work in AI is very much a boot-strapping operation and little progress can be made here unless the latest languages (such as PLANNER and CONNIVER) are readily available for use. The most suitable existing machine is a PDP 10 and no group in Britain is likely to get far without access to such a machine.

6.4 Generous support should be given over a 5 to 7 year period to any individual in Britain who can be identified as capable of making a real scientific contribution to the central area. Such support would involve provision for bringing over American workers, provision of a PDP 10, and an ample supply of studentships since it is essential to train more young people. Only if area B is supported in this way will there be sufficient trained people to make an impact on areas A and C.

6.5 If it were felt desirable to make a special effort to promote work in Basic AI, it might be worth re-examining the way such work is funded by SRC. On the definitions given here of area B, it is not really appropriate for the work to be supported through the engineering board. As already pointed out, work in AI has very close links with formal linguistics, the aim of which is to characterise formally the knowledge we have of a natural language. Formal linguistics is also in an anomalous position within the structure of SRC. AI work also has close links with psychological work on cognition and on psycholinguistics. Such work is at present funded through the Biological Sciences Committee and again the position is slightly anomalous in that the ways of thinking in this area of psychology have little in common with conventional biology. It might therefore be worth considering setting up a panel reporting direct to the Science Board to fund all three areas: the panel would take responsibility for all work aimed at gaining a better understanding of cognitive processes at a formal level. The enzyme panel forms a precedent to this idea.

Part III: Comments on the Lighthill Report and the Sutherland Reply

Dr R M Needham

1 Comment by Dr. R. M. Needham

Computer Laboratory, University of Cambridge.

Like all classifications, Lighthill's division of AI into three main parts is contentious in detail as doubtless was Caesar's similar dissection of Gaul. It would not be useful to discuss whether particular individual activities are best placed in A, B or C - at any rate if one accepts as I do the spirit of the classification. Since I basically agree with Lighthill's conclusions there is perhaps less to say than in Sutherland's commination.

The aim of the category A work is technological. Any method which achieves the desired result will do, provided it is not too expensive. This is by no means to say that it need not be founded on detailed knowledge of the subject matter, nor that it should eschew devices such as the use of heuristic methods which are perhaps associated with AI rather than with automation. On the contrary, the use of reasonably reliable heuristics is very suitable to the "no holds barred" approach. Heuristics, in general, are devices to avoid excessive searching by acting on guesses as to where to look - guesses which are not provably correct but which usually lead to something sensible. For example [To find a letter from the SRC in the Departmental office, look in the file marked SRC](#) , or [To proceed from Cambridge to Edinburgh, go first to London](#). Neither of these is always reliable, but both are based on a sound knowledge of the relevant facts.

It could perhaps be said that in any knowledge-based system some of the access rules will be formalisable and embodied in regular programs, and some will not and will thus have to be treated as heuristics. At this point, a question arises which Lighthill does not treat very much. Are there any general principles - that is, principles which apply to numerous applications - which guide or might guide the application of heuristics? Workers whom Lighthill might describe (? stigmatise) as being in category B say that such principles are, inter alia, what they are looking for. Some of the justification for this kind of work would seem stronger to those who believe that such general principles are there to be found. Lighthill suspects that they are not, so it should be pointed out that the Hart-Nilsson-Raphael theorem is one such that has been found. I do not personally think there is much to dig for here, but one should not deny that there is anything at all.

Lighthill's category C is quite outside my own technical knowledge. It is self-evident that enquiries, computer-based or otherwise, into how people work constitute an important field of scientific endeavour. The present question is about the intelligent behaviour of people, or the way people function when behaving intelligently, and whether any work can be important to this which does not explicitly concern itself with its subject matter. Which brings us to category B.

Category B work is viewed unenthusiastically by Lighthill, and defended with vigour by others. One line of defence is to call attention to developments in programming technology which it has stimulated, and to other insights to which it has led. In any venture into the history of ideas one is on dangerous ground, but in considering this kind of argument the risk has to be taken. I do not believe that the case can be made by considering programming technology. Structured programming has no dependence on AI, and the handling of complex low-level operations in terms of smaller numbers of higher-level notions has been taken to its highest development by people whose view of AI is no more favourable than Lighthill's. Backtracking is a programming technique of much antiquity. The embodiment of knowledge in procedures is a year or two younger than the act of programming; its descriptions for the

plain man is that, when looking up a table you sometimes find the address of a program to compute the value you want rather than directly being given the value itself. List-processing is a technique for burying store-management problems, excellent for rich people with complicated programs to write.

It is beyond contention that AI research has led to a great deal of excellent work in packaging these techniques attractively and embodying them in programming languages (some of which, for example, LISP, are of much interest as languages); it is a standard progression for frequently used facilities to start as library routines and end as language features.

However, the ideas did not all originate in AI, any more than did the content (though perhaps not the phraseology) of the maxim [heterarchy not hierarchy](#). The general inapplicability of strict hierarchical models, despite their seductive clarity has not merely been known but explicitly recognised as an important point by many people for a long time. To be explicit about a few aspects in my own experience: Library Automation - middle 1950's; Taxonomy - early 1960's; computer filing systems - middle 1960's. To the small extent that so vague and general a maxim can be said to be a discovery, it is one to which AI has contributed little.

Professor Sutherland remarks, in the course of a defence of category B work along these lines, that [one recent insight derived from Basic research on AI is that in interpreting the meaning of any complex input, it is impossible to use a rigid step-by-step procedure](#). Leaving aside the perhaps captious comment that this means that computers cannot do the job at all, it is emphatically not a recent insight that you cannot finish with the syntax before starting the semantics. The present writer first encountered it, in a computational linguistic context, sixteen years ago when it was not new.

This comment on a point to do with language processing leads naturally to others. Lighthill cites the understanding of natural text as one of the prime examples of the combinational explosion, and so it is. He, and also Sutherland, do in my opinion underestimate the contributions which have come from activities which are (or were not) called AI. Most people writing on such subjects tend to dismiss Machine Translation not only as a technological failure, which it was, but as an intellectually totally negligible activity, which it was not. The emphasis on exact algorithms rather than vague descriptions, the above-mentioned importance of mixing syntax with semantics, the use of heuristic devices to shorten searches in great collections of semantic data, were all studied and recognised as important. Linguistics proper, without technological objectives, has made vast progress in schematisation, exactness of description, and theoretical understanding. General interest in linguistics and in particular in rule-based (algorithmic) approaches to it, have contributed much to the intellectual climate in which AI work is done, though not always in a directly recognised manner. It is a great mistake to extract those products of other enquiries which have been found helpful by AI workers and suppose that they are AI inventions.

In sum, I do not believe that one can justify category B work by its side effects. What of its main thrust? Whether there is any middle ground between studying intelligent behaviour as an attribute of people or animals on the one hand, and making machines do complicated and useful things which used to need people, on the other, it can easily lead to a sterile philosophical debate unless we say [yes](#), on the grounds that we can seek to make machines do complicated and useless things.

[Artificial Intelligence](#) is a rather pernicious label to attach to a very mixed bunch of activities, and one could argue that the sooner we forget it the better. It would be disastrous to conclude that AI was a Bad Thing and should not be supported, and it would be disastrous to conclude that it was a Good Thing and should have privileged access to the money tap. The former would tend to penalise well-based efforts to make computers do complicated things which had not been programmed before, and the latter would be a great waste of

resources. AI does not refer to anything definite enough to have a coherent policy about in this way.

A final caution: like the majority of contributors to this paper symposium on AI, I am not an expert in any of the activities which come under its rather ill-defined umbrella. Amongst the many features, good and bad, which AI shares with Machine Translation is the fact that non-practitioners have strong views about it.

Comments on the Lighthill Report and the Sutherland Reply

Professor H C Longuet-Higgins, FRS

2 Comment by Professor H. C. Longuet-Higgins, FRS.

Theoretical Psychology Unit, University of Edinburgh.

To my mind Sir James Lighthill's most valuable contribution to the current debate on artificial intelligence has been to raise searching questions about the proper justification of the subject. We should, he suggests, ask about any piece of work whether its primary objectives are technological or scientific. If technological, such as the automatic exploration of the planets or the mechanical translation of Chinese into English, are such aims realistic in relation to our present knowledge and justifiable in economic terms? If scientific, then what science or sciences are likely to be enriched?

The need to ask such questions becomes only too apparent when one studies certain recent pronouncements on the subject (or subjects?) of artificial intelligence/machine intelligence. In the Computing Science Review recently published by the SRC the aims of [machine intelligence](#) are seen as bluntly technological, though the pious hope of achieving them [by the formulation of general principles](#) makes a hasty genuflection to scientific respectability. There is no further mention of any such principles either in the main body of the review or in the appended report of the Long Range Research Panel; all we find is an unlikely assortment of subjects grouped together under the heading [machine intelligence](#) for no better reason than [none of \[them\] seem to the Panel to demand the study of human thought or perception](#).

The subjects in question are: computational logic, real scene analysis, picture processing, [the use of the robot as an analytical tool](#), and [the acquisition of organised information in computers and interpretation of descriptive material](#). It is highly dubious whether either computational logic or real scene analysis is likely to get anywhere without due attention to our processes of thought and perception; but in any case, if such a negative criterion is adopted for what counts as [machine intelligence](#), it is difficult to see why that subject should exclude analytical geometry or analytical chemistry, which have at least as good a claim to be regarded as [analytical tools](#).

The poverty of such arguments for regarding machine intelligence as a priority area in computing science must have become plain to Sir James as soon as he undertook his penetrating survey of the field. Insofar as machine intelligence projects are basically technological, should they not be judged by the same criteria as one would apply to any piece of development work in advanced automation? Of any such project one should ask: first, what exactly is it intended to achieve, secondly, what material resources would it demand, and thirdly, what are its chances of success? Lighthill's shrewd and comprehensive critique of the technological achievements and ambitions of artificial intelligence needs no recapitulation, nor does his scepticism about the defensibility of robotics as a technological enterprise. It is only when one looks at the scientific case for artificial intelligence studies that differences of opinion seem to arise.

Sir James places in his category [C](#) all the artificial intelligence work which he regards as scientifically promising, and refers to this category as [Computer based studies of the central nervous system](#). In so doing he aligns himself with those of us who hold that the main justification for artificial intelligence is the light it can throw upon human intellectual activity. But his chosen heading, and some of his later remarks, indicate that he attaches more significance to work on the [hardware](#) of the brain than to work on its [software](#).

This is the only point on which I want to take issue with him. He is, of course, perfectly right in saying that anyone who is developing network models of the brain had better work within the constraints imposed by our knowledge of its anatomy and physiology; it would be foolish for an engineer to speculate about the circuitry of a computer when he could perfectly well open it up and look inside.

But the hardware of computers is very far from being the only matter relevant to their functioning. In order to understand how a computing system works one must enquire into the logic of the system software and the semantics of the programming languages in which the system can be addressed. The corresponding questions about human beings are those asked by the science of psychology - though admittedly, psychological theories seldom attain a degree of sophistication worthy of their subject matter. An outstanding exception to this stricture is the science of linguistics, and perhaps it is no coincidence that the most impressive achievement of artificial intelligence to date is a working model of the comprehension of natural language.

I would go further and hazard the prediction that for some time to come the most valuable work in artificial intelligence will be that which attempts to express, in the form of computer programs, abstract theories of our various cognitive faculties, rather than mathematical models of the brain itself - this in spite of some excellent recent work on the possible role of the neocortex as a classifying device. This view is based not only on the obvious vitality of current artificial intelligence work on language and vision, but also on an evident dissatisfaction among psychologists with the naive stimulus-response theory of behaviour as it has been applied to human beings.

It is now plain that a central problem in cognitive psychology is to understand how our knowledge is represented and deployed, and the computer program is the only medium which at present offers us the possibility of formulating adequately sophisticated theories of cognition. The elimination of inadequate theories is no longer the main problem; the defects of a programmed theory become immediately apparent as soon as it is run on a computer.

In short whatever the technological prospects of artificial intelligence, its principal scientific value, in my view, is that it sets new standards of precision and detail in the formulation of models of cognitive processes, these models being open to direct and immediate test.

The question [What science or sciences are likely to be enriched by artificial intelligence studies?](#) can now receive a provisional answer, namely [All those sciences which are directly relevant to human thought and perception](#). These [cognitive sciences](#) may be roughly grouped under four main headings:

1. [Mathematical](#) - including formal logic, the theory of programs and programming languages, the mathematical theory of classification and of complex data structures.
2. [Linguistic](#) - including semantics, syntax, phonology and phonetics.
3. [Psychological](#) - including the psychology of vision, hearing and touch, and
4. [Physiological](#) - including sensory physiology and the detailed study of the various organs of the brain.

Perhaps [cognitive science](#) in the singular would be preferable to the plural form, in view of the ultimate impossibility of viewing any of these subjects in isolation. Indeed artificial intelligence studies are beginning to offer interesting suggestions as to how our various modes of experience might be logically related.

Finally, perhaps one should say a word about the main point of disagreement between Lighthill and Sutherland. Professor Sutherland's redefinition and reinstatement - of Lighthill's

category B as [basic artificial intelligence](#) has my sympathy, because although I hold no particular brief for [bridging activities](#) as such, I do think that there is a place in artificial intelligence for studies which are addressed to the general problems which have been found to recur in many different areas of cognitive science. The mathematician's ability to discover a theorem, the formulation of a strategy in master chess, the interpretation of a visual field as a landscape with three cows and a cottage, the feat of hearing what someone says at a cocktail party and the triumph of reading one's aunt's handwriting, all seem to involve the same general skill, namely the ability to integrate in a flash a wide range of knowledge and experience. Perhaps Advanced Automation will indeed go its own sweet way, regardless of Cognitive Science; but if it does so, I fear that the resulting [spin-off](#) is more than likely to inflict multiple injuries on human society.

Comments on the Lighthill Report and the Sutherland Reply

Professor D Michie

Department of Machine Intelligence and Perception, Edinburgh.

Two contrary attitudes are common. In the first place there is a widespread, although mostly unconscious, desire to believe that a machine can be something more than a machine, and it is to this unconscious urge that the newspaper articles and headlines about mechanical brains appeal.

On the other hand, many people passionately deny that machines can ever think. They often hold this view so strongly that they are led to attack designers of high-speed automatic computing machines, quite unjustly, for making claims, which they do not in fact make, that their machines have human attributes.

M. V. Wilkes 'Can a machine think?' in *Discovery* May, 1953

Sir James Lighthill's report speaks of [the ABC of the subject](#), categorising it as follows:

- A - Advanced Automation
- B - Building Robots
- C - Computer-based CNS research

The report regards [A](#) and [C](#) as worthy activities which, however, have made disappointing progress. [B](#) is regarded as unworthy, and as having made very disappointing progress indeed. [B](#), it should be noted, is really used in the report to denote any experimental programming which lacks obvious application to either [A](#) or [C](#). Thus computer chess is included in [B](#) whereas robot parcel-packing is put into [A](#).

Most people in AI who have read the report have had the feeling that the above classification is misleading. Sir James has arrived at his position by interpreting AI as consisting merely of outgrowths from a number of established areas, viz.:

- A as an outgrowth from control theory,
- B as an outgrowth from science fiction,
- C as an outgrowth from neurobiology,

These interpretations are remote from those current in the field itself.

A number of questions accordingly pose themselves, including the following:-

1. Was this report based on as thorough a survey as it should have been? In particular, was opportunity taken to invite the views of the international leaders of the field?
2. How successful has the author been in overcoming the difficulties inherent in his inexperience of the field, and in putting aside his own professional biases?
3. Has accepted practise been followed in documenting subjective opinions wherever possible and in providing factual sources and references which others can check?
4. What is the validity of the [A B C](#) classification? Would the computing science community accept it?
5. Are the report's assessments of work in the [B](#) category - [Building Robots](#) - intended to apply to experimental robotics conducted in the United Kingdom? If so, should not the author
 - i. have said so plainly,
 - ii. have asked to see the experimental robotics work during his visit to Edinburgh?

International opinion not consulted

The first of these questions is so critical as to merit a brief note to the effect that the leading American workers, such as McCarthy, Minsky, Nilsson, Raphael and Robinson, were not in fact consulted. The appearance of their names in the list of fifty given in the report's third paragraph derives from the fact that the author has read scientific writings of theirs not that he invited their opinions. Since the field in question was pioneered in the United States of America, which supports to this day an effort on at least twenty times the scale of that in the United Kingdom, it is well to bear in mind this fact when assessing Sir James' evaluations.

Intelligence Theory

Space does not allow a review of the remaining questions of the above list; a detailed critique is available elsewhere. (D. Michie (1972) [On first looking into Lighthill's Artificial Intelligence report](#) (unpublished)) Instead I will briefly indicate two themes which arise from Lighthill's implicit question: [If you throw A and C away, what is left, if anything?](#) Lighthill's answer is [Building Robots](#). An alternative answer, which many will prefer, is [Intelligence Theory](#). By this we mean attempts to systematise the design principles of intelligent systems wherever they may be found, whether in the [A](#) or [C](#) application areas.

Having fixed on [Building robots](#), Lighthill paints a picture of this pursuit which must strike those actually engaged in experimental robotics as somewhat unfamiliar. In studies of actual robot work the role of the equipment is plainly seen as test gear for putting certain types of theoretical ideas to experimental test. A pertinent parallel is the building of wind-tunnels as an aid to aero-engineering - as illustrated in the figure below.

Broad subdivisions of two fields of enquiry according to theoretical-experimental and technological-biological classifications.

This [figure](#) brings into relief the reason why [Building Robots](#) is an unhelpful choice for the role of Bridge between [A](#) and [C](#). It is surely more fruitful, if one seeks inter-disciplinary connections, to choose a common body of theory rather than to seize on a piece of laboratory equipment. One feels that Sir James would be among the first to agree that to speak of [Building Wind-tunnels](#) as the Bridge between aero-engineering and the study of bird flight, would direct attention away from the true bridge, namely the science of aerodynamics. The equivalent science in the case of AI is at a primitive stage. It is the hope of every AI professional to contribute in some way to bringing the required theory into being. This, as I see it, is the burden of Sutherland's re-definition, in his contribution to this symposium, of [B](#) as standing for [Basic](#).

Encouragement of research in machine intelligence

On this note I would like to leave Sir James Lighthill's interesting and imaginative review and to mention an assessment of a more home-spun quality: the report of SRC Computing Science Committee's long-range panel, published in [Computing Science Review](#). This panel, composed of computer professionals, considered the machine-oriented part of Artificial Intelligence (ie the [A+B](#) part) and recommended that special encouragement should be given to this field. However, it is evident to those who work in the field that it would be helpful if a clear and concise statement were given of its goals and methodology. The style of Sir James Lighthill's report suggests that there is a lack of understanding in some quarters, and without this there is a reluctance to recommend significant expenditure. The status and position of the subject are particularly clear at the moment and it is, therefore, opportune that a statement should be made to avoid any further misunderstandings.

The subject, in so far as it comes within the Computing Science Committee's realm of interest, is concerned with machines, and in particular computers, displaying characteristics which would be identified in a human being as intelligent behaviour. Perhaps the characteristics which are most important are those of learning and problem solving. The

applied benefits which may be gained from work in this field could bring considerable economic benefit to the country. They are two-fold:

- a. To relieve the burden at present on the systems analyst and programmer in implementing applications;
- b. To enable new and more complex applications to be undertaken in this country in competition with work elsewhere.

These are the long-term advantages and to this end work is proceeding on a number of detailed problems, including the following:

- I. Automatic assembly and other robotic applications
Mass spectogram analysis
Chemical synthesis planning
Assembly line balancing
- II. Language-understanding systems
Semi-automatic programming (ie [teachable systems](#)) and ultra-high level programming languages (like PLANNER, SAIL, CONNIVER)

Group I represents useful applications. Group II represents the subject's own special contribution, independent of specific applications to computer science. This lies in making it more possible for the user to get computing systems to [understand](#) what he means.

Many good scientists have been involved in this field and their work has resulted in the development of techniques and methods of wider use, for example:

- List-processing was originally devised by Newell, Shaw and Simon for AI work and first implemented in their IPL language.
- The incorporation of conditional expressions into ALGOL 60 was McCarthy's suggestion derived from his work on LISP, itself inspired by the needs of Artificial Intelligence work.
- The POP-2 language, now implemented on 5 main hardware ranges was specifically developed for AI work, but subsequently shown to be of wider utility.
- In fifteen years of struggle towards [language understanding](#), striking advances have been scored (Bobrow, Winograd, Woods).
- Some of the search and associative techniques used by programmers and operations research workers have been initiated in AI, and assimilated without awareness of their origin.

The problems that have been mentioned above are practical problems. Abstracting from these, and observing the methods of solution, workers in the field have been able to define general principles for intelligent systems. This work has made some progress and the following theorems and methods have been developed.

[Some of these are reviewed in a Nature article \[Machines and the theory of intelligence\]\(#\) 23 Feb. 1973](#)

>Problem-solving

- Theorems of minimality and completeness of various algorithms for heuristically guided search.
- Methods of pruning search trees in special situations: Plausibility analysis; alpha-beta pruning.
- Recursive formation of sub-problems as in Newell and Simon's [General Problem Solver](#).
- Application of theorem-proving ideas in problem-solving.
- Studies of problem-representation.

Recognition

- Various methods of feature extraction and interpretation for visual data.
- Use of semantics to disambiguate linguistic analysis.
- Matching of descriptions represented as directed graphs (eg [hierarchical synthesis](#)).

Learning

- Adaptive learning via parameter-optimisation.
- Rote-learning techniques.
- Formation of new concepts from examples and counter-examples.
- Inductive generalisation.

Even so incomplete a list as the above puts into perspective the importance of examining particular problems in depth (such as chess-playing or those involving robots) so as to investigate how to bring the above functions to bear in an integrated fashion. They are but experiments which may be used to derive or test theories. At this early stage of innovation the overwhelming benefit to be derived from a given experimental study lies in its role as a forcing function for new programming techniques and tools. The field is so difficult and the choice of the right problem at the right moment so much part of the art of enquiry that this should be left to the research workers themselves. They should be judged by their success or otherwise in advancing the state of computer programming, and in introducing and testing computer languages of greater expressive power.

Footnote on Sutherland's commentary

Sutherland's otherwise admirable analysis contains two expressions of view with which exception must be taken, namely (1) that AI should not be handled by the Engineering Board of SRC and (2) that AI research in Britain is in a bad way.

1. A reasonable approach would surely be to distinguish A-oriented and C-oriented poles of the subject and to provide for the first under the Engineering Board and for the second under the Science Board. Since important contributions continue to be made by computer scientists ignorant of psychology and brain science, and by psychologists ignorant of computer science, it would avoid embarrassment, and reflect scientific reality, to make separate provision.
2. Sutherland's proposition that AI research in Britain is in a bad way deserves to be vigorously challenged. But it is inappropriate for me, as founder of the longest-established British research group, to be the one to do this. A better corrective can be obtained from assessments by authoritative outside observers, such as that by Dr. Nils Nilsson (*) of the Stanford Research Institute's Artificial Intelligence Centre and author of the graduate textbook [The problem-solving methods in Artificial Intelligence](#).

* [An outsider's view of the Experimental Programming Unit at Edinburgh University, obtainable from School of Artificial Intelligence, Edinburgh.](#)

What is to be done

We are in the embarrassing situation in Britain that in order to carry out significant work over the next few years in the context of international competition, it will be essential to import American machines - specifically the DEC System 10 (formerly known as the PDP 10). I think that everybody would be happier about the case for allowing an American importation now if steps were at the same time taken [to see that British AI research never found itself in such a predicament again](#). What would have to be done if this desirable state of affairs were to be brought about?

Why does the need arise? It is not only, or primarily, because of the superiority of the architecture of the DEC System 10 for AI-type uses. The over-riding consideration is access to the rapidly accumulating fund of AI-oriented software and applications programs in the big American laboratories. The key to the situation is the absence of software for British machines, either present or **new range**, suitable for AI work, which has its own very peculiar needs. These needs are peculiar. It is hardly more sensible to speak of **making do** with, say, general scientific software developed without reference to AI than to suggest that, say, plasma physicists short of experimental fusion equipment should **make do** by borrowing linear accelerators from the particle physicists!

The kind of software development needed if AI workers are ultimately to be put in business as users of the new range of British computers (I do not necessarily intend this phrase to be exclusively confined to ICL) comes under two headings:

1. Development of experimental operating systems, compilers and packages, as has been done in a small way on the ICL 4130 at Edinburgh. But the new effort should aim to embrace the entire **standard range** of facilities which every AI worker should be entitled to take for granted LISP, POP-2, SNOBOL, QA4, PLANNER, CONNIVER, etc. etc., and, ultimately far more important, **leap-frogging** into the future both by adapting the latest advances of AI research work where appropriate, and by innovation within the R & D effort itself. Also to be considered are operating system features for handling **funny peripherals** (experimental robots, speech input devices, etc.) and basic packages for **front end** functions such as, say, video and speech input, robot control functions, language pre-processing. In addition the design and development of advanced peripherals (eg for robotics) should be regarded as an integral part of the job although (as with software) the more standard aspects of instrumentation should be contracted out to industry wherever possible.
2. **Community aids** whereby new research programs and software developed in overseas laboratories can be made immediately available on demand for British research workers to test out and either accept or reject as tools for their own needs. Community can be achieved by various means and these means will vary according to the nature of the case, but they include software/hardware interfaces to the ARPA net, and emulation (for example by microprogramming) of the 'donor' machine from which the program is to be adapted.

If a fully-fledged development project is to be got up to full speed by around 1977 then forward studies could usefully be started now. It is already obvious that early installations of a PDP-10 in an active centre of British AI research is a precondition if these studies are to develop fruitfully, since immediate access to the latest AI research materials (and intimate contact with advanced AI research) will be as essential to the specification and development of new research facilities as it is for those who will later be using them. Until the new facilities exist, the only point of access and contact will be through British groups equipped compatibly with their American counterparts, - ie with PDP-10s.

In this field successive workers in a given area should be able to stand on the shoulders of their predecessors through the medium of successive contributions to a common stock of new language aids and library packages. This will not happen unless someone makes it his business continually to scoop in what is new and useful and build it into a properly documented and integrated system. The level at which the British AI community will be able to contribute in the late 1970s, as judged by competitive international standards, will be crucially affected by the sophistication of the available software.