Part II Some comments on the Lighthill Report and on Artificial Intelligence

what's he done

N. S. Sutherland, Professor of Experimental Psychology, University of Sussex. *August 1972.*

Introduction

No one is likely to quarrel with Lighthill's clasification 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 neuro-physiological 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

On page 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 (pages 7 and 8) 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.

The question boils down to the existence of such principles

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.

example?

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.

when does This fit

ACCEPT

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

but it may cloud his judgement of its success 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.)

This is the canonical area B result—but maybe introspection would work as well

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 *succès d'estime* of Winograd's program partly because their authors were not so adept at putting their ideas across in a readily understandable way.

I don't believe Winograd's work was necessarily dependent on planner

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 practioners 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.

show a questionable lack of detachment

not clear how about graduate students like me

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.

Should give how rebutted claim + need for limitations

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 lead 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.

They are ad hoc tricks, but that's ok

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.

They better be good to go from nl. to non?

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 any one 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

LISP isn't exactly the ideal language for doing this

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.

no evidence that the languages are necessary

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.

rather strong language for such meager progress

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.

all negative

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.

doesn't it work as well

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 lead 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.

what was the actual origin

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.

how

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.

[illegible] the difference between being a student [illegible]

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.

B

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 post-graduate 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.