Probably the ultimate achievement in simulating human operation is the creation of an artificial intelligence. At present, artificial-intelligence research has revealed aspects of human ignorance, particularly in the fields of linguistics and psychology. In the future, it is desired to produce machines which can translate languages, recognise printed letters and, most important of all, possess a capacity for self improvement.

The search for artificial intelligence

by **R. J. SOLOMONOFF**, M.Sc.

REASONS for interest in artificial intelligence vary considerably. Certain approaches to various problems involve attempts to simulate man's solution of them. This work amounts to a renaissance of introspective psychology. In the past, an introspective psychologist would write a description of how he felt that his own mind worked. There was at that time no method of checking the accuracy of such conjectures.

At the present time, if a psychologist writes a description of how he feels that he solves problems in a certain field, he can now partially check his description from two aspects. First, does his description describe anything with adequate specificity? If it does, then he can write a computer program to simulate it, but, if it does not, the psychologist must try to fill in details that he has left out; his understanding of the process is certainly inadequate.

If his description is adequate, he can then use the resultant computer program in attempts to solve problems in the field of interest. While failure would be strong evidence of the inadequacy of the original introspective conjecture, success in working these problems can only be regarded as partial corroboration. Further corroboration can be obtained if only slight modifications of the original program are needed to work problems in very different fields.

There are some workers in the artificial-intelligence field who are only mildly interested in the operation of the human

A good approach to a definition of artificial intelligence is that of Fein,* who lists 14 types of problems that people in this field are building machines or writing computer programs to solve. These include devices to:

- **comprehend live voice and handwritten messages**
- **prove theorems in mathematics and logic**
- **identify targets and recognise subatomic particles from bubble-chamber tracks**
- **form concepts, generate hypotheses and make inductive inferences**
- **translate printed French, English, Russian, German, Chinese and Japanese**
- **predict weather, make medical or psychiatric prognoses, explore space, explain the space environment, and make spatial predictions**
- **act as a reference librarian**
- **act as a mathematical assistant**
- **prescribe medical treatment**
- **write and debug computer programs**
- **design intelligent machines**
- **memorise data and answer questions**
- **play games with incomplete strategies**
- **imitate or emulate a military commander, a government head, a business corporation and a nation's economy.**

While the list is by no means exhaustive, it gives a good idea of what we mean by 'artificial intelligence', and might be regarded as an inductive definition.

More generally, we would like the machines to learn to improve their operation after working on a set of problems. While superficially most of these problems seem to differ much from one another, many workers in this field (including the author) believe that they have much in common and that progress in one of the problems can often be translated into progress in several of the others.

* **FEIN, L.: 'The artificial intelligencia',** ***IEEE Spectrum*****, Feb. 1964, pp. 76–87**

If $I=\int_0^\infty exp(-x^2)\,dx$

then it follows that

$I=\int_0^\infty exp(-y^2)\,dy$

and that

$I^2=\int_0^\infty exp(-x^2)\,dx\int_0^\infty exp(-y^2)\,dy$

$=\int_0^\infty\int_0^\infty exp(-x^2)\,exp(-y^2)\,dx\,dy$

Transforming to polar co-ordinates

$(r^2=x^2+y^2,\ dx\,dy=r\,dr\,d\theta)$

$exp(-x^2)\,exp(-y^2)=exp-(x^2+y^2)$

$=exp(-r^2)$

Therefore $I^2=\int_0^\infty\int_0^{\pi/2} r\,exp(-r^2)\,dr\,d\theta$

$=\int_0^\infty r\,exp(-r^2)\,dr\int_0^{\frac{\pi}{2}}d\theta=\left[-\frac{1}{2}exp(-r^2)\right]_0^\infty\left[\theta\right]_0^{\frac{\pi}{2}}$

Therefore

$I^2=-\frac{1}{2}(0-1)\left(\frac{\pi}{2}\right)=\frac{\pi}{4}$

and $I=\pm\sqrt{\frac{\pi}{4}}$

Therefore $\int_0^\infty exp(-x^2)\,dx=\sqrt{\frac{\pi}{4}}$

taking the positive value

mind. They use introspection to obtain clues on how to write clever computer programs, but, if they think of a very effective program that seems to have no correspondence in their own mental processes, they will not hesitate to use this program in solving problems.

Simulation

While the author's interest is mainly in obtaining machines to solve problems in a very effective manner, for the preliminary part of the work it is expedient to try to simulate human problem solution as closely as possible, since it is more likely that we will be able to find and correct errors in a device that works somewhat the way that people do.

In general, one tends to think about these machine programs in ways that model one's own thought. These ways of thinking will tend to be more correct if the machines do, indeed, operate much the same way as people do. After proceeding as far as possible in the machine simulation of human-thought processes, modifications can be written of these programs that take advantage of the technological peculiarities of nonbiological computers.

While each area of artificial-intelligence work usually has its own peculiar motivation—sometimes immediately practicable, sometimes purely scientific—the understanding of human-thought processes is a thread that seems to pass through most of them.

Expectation

What can be expected from this research? If the goal of simulating many of man's higher-order creative problem-solving skills is achieved, important rapid cultural changes are assured. If man's skills can be equalled, it is fairly certain that they can be exceeded, in speed at least, and also in the availability of rapidly accessible memory.

An immediate result would be a very significant increase in the rate of development of science and technology. Since man, at the present time, has difficulty in coping with the present rate of technological development, it can only be hoped that the very intelligent machines will be of some help in solving the problems that they will create.

Suppose this goal is never achieved—or that we must wait a thousand years for it to be achieved—what developments are now available from artificial-intelligence research, and what can be expected in the near future? The principal past benefit is that this research has revealed the extent and nature of human ignorance in several fields that were thought to be understood. In linguistics and psychology, this revelation has been most noticeable.

Linguistics

In linguistics, the lack of understanding of grammar and semantics was made clear when feeble early attempts were made to write computer programs to translate from one human language to another. This unsuccessful beginning spurred a rebirth of linguistic science—entirely new ideas on how languages are constructed and how their structures may be related to meaning. Although an adequate grammar for any human language has not yet been devised, and the science of semantics (the relation of sentence structure to its meaning) is in a very rudimentary state, some idea is now known as to the extent of human ignorance, and a criterion is available for determining when linguistics is understood. Promising new starts have been made in these areas.

In psychology, as has been mentioned, there has been a rebirth of introspection as a valid method of investigation—provided that its conjectures are properly checked by simulation.

Practical outgrowths

At the present time, there are few practical outgrowths from artificial-intelligence work. One of these is a program for translating from Russian to English language. The translations are not very good, but are adequate for certain work. The machine is faster than human translators, but, at present, more expensive. It is fairly certain that the cost will improve in the near future, following the decreasing cost of computer use.

Another practical outgrowth is the use of machines for recognising printed letters. At the present time, they are quite capable of doing the task if only one fixed alphabet is used. There are under development various recognisers for less-carefully-printed characters or hand-lettered characters, but they are less reliable. Also, there is a program for playing draughts that will probably beat most people who are not close to being world champions.

It seems clear that there are not many immediate practical benefits of research in this field. It is therefore reasonable to discuss how research is progressing and which are the most promising developments.

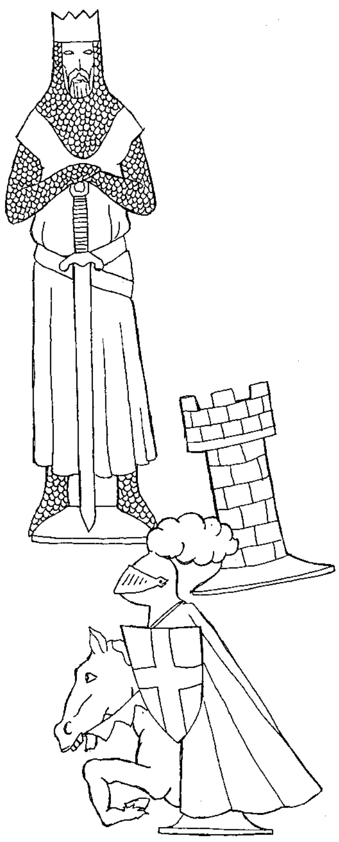
Research

In a field where developments are rather rudimentary, it is the author's belief that the course of research most likely to terminate in very intelligent programs is in devising machines that have the capacity for self improvement, as well as for solving other problems normally requiring human intelligence.

Almost all of the artificial-intelligence work on problems of sufficient complexity to deal with self improvement has been in the area called heuristic programming. Here, researchers devise methods to solve problems that do not necessarily always work, but are successful frequently enough to warrant trying them under the right conditions. Those methods are usually obtained by introspection; the experimenter is modelling part of his own mind within the machine.

Heuristic program

One of the best known heuristic programs is called GPS—general problem solver. It was devised to work on a broad class of problems the way humans seem to. Suppose someone is in New York and the goal is to go to Rio de Janeiro. Immediately, there are three possible



general ways: by sea, land or air. Each of these alternatives in turn has many sub-alternatives; e.g. to go by sea one might go by commercial steamship or buy a sailboat of one's own. Each of the alternatives considered might be limited by various constraints, e.g. the amount of time or money available for the trip. These limitations can be taken as fixed, or regarded as new subgoals, e.g. getting enough money to buy a boat might be such a subgoal.

As the original problem of going to Rio is investigated, a large network of goals, alternative subgoals and alternative subsubgoals emerges. Some of the subgoals can be achieved only by simulataneously satisfying several subsubgoals. If one decides to buy a boat and sail to Rio, both the necessary money and the sailing skill must first be obtained.

In other cases, it is not certain whether certain subgoals will be effective in attaining their main goal—one can only assign probabilities. For instance, the sailing is done by an inexperienced person, a probability of 0·1 of arriving in Rio intact may have to be assigned.

Tasks

This analysis of the original problem results in a very large treelike structure of interrelated tasks. On making various simplifying assumptions about this network of possible tasks, GPS was often able to choose a sequence of goals and subgoals that solved the original problem.

Superficially, this network of tasks resembles a conventional PERT network, but the fact that completion of many of the tasks is uncertain makes the nature of the problem very different. Relatively recently, there has been work on PERT networks with uncertainties in the costs or completion times.

GPS was certainly an important generalisation in artificial-intelligence work. For each type of problem, special routines had to be devised to find subgoals and evaluate them before the GPS technique could be applied. The simplifying assumptions used were rather extreme, and the model of human problem solution used was a very rudimentary one. It did, however, solve problems in logic and did so in ways that seemed similar to the ways humans solved them. While GPS undoubtedly had many deficiencies, it is the author's view that it is the hypothetical 'talking dog'. One does not complain that the dog does not deliver well modulated Shakespearian prose—that he talks at all is a miracle!

Following and improving upon GPS, there have been many good problem-solving programs. Proving theorems in geometry and formal integration of symbolic expressions were some of the early successes. While GPS was originally designed to try to simulate human problem solution, the most effective developments following it have tended to diverge from human-like methods.

Learning ability

One of GPS's deficiencies was its very poor learning ability. If similar problems were successively presented to it, GPS could not ordinarily use the known solutions of earlier problems to shorten the solution times of later problems. Slagle, who wrote one of the early successful formal integration programs, has devised a problem solver called *Multiple* that modifies its estimates of the probability of resolution of various tasks on the basis of its past experience with simlar tasks.

Hormann has a problem solver called GAKU that has more general learning capabilities. She has done much work on the problem of 'secondary learning'; in addition to learning from its own experience, it should be possible for an instructor to 'tell' the machine various generalisations derived from the experiences of the instructor. At present GAKU is probably the most general problem solver that has much learning ability.

Difficulties

What are some of the important difficulties in heuristic programming? The task-net problem is certainly one. Given a network of tasks and subtasks that is related to the solution of a problem with the probabilities of success and expected time to solve each subtask, which task should one work on next? Slagle has solved a simplified form of this problem, but has not yet applied his solution to heuristic programming.

The problem of induction is of much importance. Given all the data that one feels are relevant to a certain question, how can one express the probabilities of the various possible answers in terms of those data? In its most general form, the solution involves devising new concepts and using them for extrapolation. For a long time, the author has felt that this problem was critical in understanding and simulating any kind of learning, and he devised a very general mathematical theory for making the best possible predictions from a given set of data. It is possible to show, however, that this theory, and any other general theory of induction satisfying certain natural requirements, cannot in principle be realised in practice.

It is, however, possible to make realisable approximations to these unrealisable ideal theories, and all practical learning can be regarded as approximations to these ideals. At present, the ideal theories are useful for criticising and correcting more practicable forms of induction.

An interesting form of induction is analogy. A program has been written to work geometric-analogy problems such as occur on IQ tests. The use of analogy in human problem solution is quite important, and this program makes the concept of analogy sufficiently precise to be used in a variety of problem-solving situations, although, so far, no further applications have been made.

Minimal machine

Two interesting kinds of induction use simple machines for extrapolation. The first method starts with a small set of sentences in Hungarian, and their corre-

аемость = ability
фталевый = phthalic
лактаза = lactase
растворимый = soluble
требуемый = required
бутан = butane
красноватый = reddish
теория = theory
анемия = anemia
флюксия = fluxion
неврит = neuritis
битый = beaten

sponding translations into Basque. A minimal machine is then devised that will translate those Hungarian sentences into the corresponding Basque. There are known routines for constructing such machines, and a 'minimal' machine is one that has the least possible number of states; it is, in a sense, least complex. This minimal machine is then given new Hungarian sentences, and it is able, in many cases, to translate them correctly.

Another way to use simple machines for induction starts with a sequence of symbols that was generated using some simple rule. Here the rule can, but need not, have some random effects in it. A randomly obtained small machine is then used to try to create this same sequence of symbols. It will get a certain fraction of them correct—its 'score'. This initial machine is then changed slightly at random (mutated), and the new machine is tried, in attempts to reproduce the initial sequence. If it gets a lower score than the previous machine, the previous machine is mutated in a different way to obtain another new trial machine. If it gets a higher score than the previous machine, the previous machine is discarded and the new machine in turn is mutated to produce a new trial machine.

Mutation

This process of mutation and 'selection of the fittest' is continued until a small machine is found that can adequately produce the original sequence. Such machines tend to continue this sequence in accord with the rule (initially unknown to the machine) by which the sequence was originally created.

The process described bears much resemblance to organic evolution, but is far simpler, both in the problems it solves and the methods it uses. If it is to solve more difficult problems, some of the more complex aspects of organic evolution will have to be simulated in the system. One aspect of this kind might be the preservation in the new trial machine of a set of features of the parent machine that seem to work well together.

Another important kind of problem undergoing interesting development involves the input and output transformations of the data.

Language

On one level there has been work on machines that understand languages similar to English. Bobrow has devised one of these in which a word problem in algebra is presented to the machine in a kind of simplified English. The machine is able to extract the necessary information from these sentences, and express it in its own internal language. It then solves the desired problem if it can and types out the answer in the same simplified English. If there is not enough information for the machine to work the problem, it may ask questions.

Amarel was one of the first to draw attention to the importance of 'the transformation problem' in a more general form. Just as Bobrow's machine first transforms the input problem to an internal form in which it can be more readily solved, so too, many problems are solved by humans by first transforming them into familiar forms in which the solver has more ideas on how to solve them. Ideally the unknown problem should be transformed into an equivalent one that the solver has already solved or is fairly certain that he knows how to solve.

The internal languages of the machines pose, in almost all devices attempted to date, severe limitations on what the machine can accomplish. These languages are limited so that there are certain concepts that they could not conceivably express. In most cases, this is certainly an unnecessary limitation. 'Universal languages' that can express anything are known. Probably all direct digital-computer languages are universal in this sense. While it is usually easy to show that a language is universal, if indeed it is, it is usually quite difficult to show that one is not universal if it is not.

The languages used by a broad class of machines called 'perceptrons' have been shown by Minsky and Papert to be incapable of expressing certain concepts, but this proof was very difficult. Ideally, a machine should not only be capable of expressing any conceivable concept, but also it should be capable of expressing commonly used concepts by simple combinations of its basic vocabulary. From a practical standpoint, if the universality of a machine language is not obvious, the machine should be modified so that its language is clearly universal.

Complex concepts

It is perhaps notable that the work in character recognition or multidimensional-pattern recognition has not been described in any detail here. In general, the concepts expressable by any of the systems proposed thus far have been very limited—certainly they do not have universal languages. As such, it is extremely unlikely that such devices can be used to generate the very complex concepts needed in linguistic processing or in the intricacies of devising programs to improve other programs.

At the present time, self-improvement capability does not appear to be an important problem to work on. It is necessary, however, to have problem solving programs as general as possible, having universal languages for expressing concepts. If these programs are able to deal effectively with problems, such as generalised induction and the writing and improvement of computer programs, the capacity for self improvement will not be a large step from this point.

Bibliography

MINSKY, M.: 'Steps toward artificial intelligence', *Proc. Inst. Radio Engrs.*, 1961, **49**, pp. 8–30

MINSKY, M.: 'Artificial intelligence', *Sci. American*, Sept. 1966, pp. 246–260

SOLOMONOFF, R. J.: 'Some recent work in artificial intelligence', *Proc. Inst. Elect. Electronics Engrs.*, 1966, **54**, pp. 1687–1697