Collected Works of A.M. Turing

MECHANICAL INTELLIGENCE

Edited by

D.C. INCE
Open University, Milton Keynes, United Kingdom

1992
NORTH-HOLLAND
AMSTERDAM · LONDON · NEW YORK · TOKYO

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
655 Avenue of the Americas
New York, NY 10010, USA

ISBN: 0 444 88058 5

Library of Congress Cataloging-in-Publication Data

Turing, Alan Mathison, 1912–1954.
Mechanical intelligence / edited by D.C. Ince.
p. cm. -- (Collected works of A.M. Turing)
Includes bibliographical references and index.
ISBN 0-444-88058-5
1. Artificial intelligence. I. Ince, D. (Darrel) II. Title.
III. Series: Turing, Alan Mathison, 1912–1954. Works. 1990.
Q335.5.T87 1992
006.3--dc20 90-36187
CIP

This book is printed on acid-free paper.

Printed in The Netherlands

Intelligent Machinery

A. M. Turing
[1912—1954]

Abstract

The possible ways in which machinery might be made to show intelligent behaviour are discussed. The analogy with the human brain is used as a guiding principle. It is pointed out that the potentialities of the human intelligence can only be realized if suitable education is provided. The investigation mainly centres round an analogous teaching process applied to machines. The idea of an unorganized machine is defined, and it is suggested that the infant human cortex is of this nature. Simple examples of such machines are given, and their education by means of rewards and punishments is discussed. In one case the education process is carried through until the organization is similar to that of an ACE.

I propose to investigate the question as to whether it is possible for machinery to show intelligent behaviour. It is usually assumed without argument that it is not possible. Common catch phrases such as 'acting like a machine', 'purely mechanical behaviour' reveal this common attitude. It is not difficult to see why such an attitude should have arisen. Some of the reasons are:

(a) An unwillingness to admit the possibility that mankind can have any rivals in intellectual power. This occurs as much amongst intellectual people as amongst others: they have more to lose. Those who admit the possibility all agree that its realization would be very disagreeable. The same situation arises in connection with the possibility of our being superseded by some other animal species. This is almost as disagreeable and its theoretical possibility is indisputable.

(b) A religious belief that any attempt to construct such machines is a sort of Promethean irreverence.

(c) The very limited character of the machinery which has been used until recent times (e.g. up to 1940). This encouraged the belief that machinery was necessarily limited to extremely straightforward, possibly even to repetitive, jobs. This attitude is very well expressed by Dorothy Sayers (*The Mind of the Maker* p. 46) '. . . which imagines that God, having created his Universe, has now screwed the cap on His pen, put His feet on the

mantelpiece and left the work to get on with itself.' This, however, rather comes into St Augustine's category of figures of speech or enigmatic sayings framed from things which do not exist at all. We simply do not know of any creation which goes on creating itself in variety when the creator has withdrawn from it. The idea is that God simply created a vast machine and has left it working until it runs down from lack of fuel. This is another of those obscure analogies, since we have no experience of machines that produce variety of their own accord; the nature of a machine is to 'do the same thing over and over again so long as it keeps going'.

(d) Recently the theorem of Gödel and related results (Gödel 1931, Church 1936, Turing 1937) have shown that if one tries to use machines for such purposes as determining the truth or falsity of mathematical theorems and one is not willing to tolerate an occasional wrong result, then any given machine will in some cases be unable to give an answer at all. On the other hand the human intelligence seems to be able to find methods of ever-increasing power for dealing with such problems 'transcending' the methods available to machines.

[[1]] (e) In so far as a machine can show intelligence this is to be regarded as nothing but a reflection of the intelligence of its creator.

REFUTATION OF SOME OBJECTIONS

In this section I propose to outline reasons why we do not need to be influenced by the above-described objections. The objections (a) and (b), being purely emotional, do not really need to be refuted. If one feels it necessary to refute them there is little to be said that could hope to prevail, though the actual production of the machines would probably have some effect. In so far then as we are influenced by such arguments we are bound to be left feeling rather uneasy about the whole project, at any rate for the present. These arguments cannot be wholly ignored, because the idea of 'intelligence' is itself emotional rather than mathematical.

The objection (c) in its crudest form is refuted at once by the actual existence of machinery (ENIAC etc.) which can go on through immense numbers (e.g. $10^{60,000}$ about for ACE) of operations without repetition, assuming no breakdown. The more subtle forms of this objection will be considered at length on pages 18–22.

The argument from Gödel's and other theorems (objection d) rests essentially on the condition that the machine must not make mistakes. But this is not a requirement for intelligence. It is related that the infant Gauss was asked at school to do the addition $15+18+21+\ldots+54$ (or something of the kind) and that he immediately wrote down 483, presumably having

[[2]] calculated it as $(15+54)(54-12)/2.3$. One can imagine circumstances where a foolish master told the child that he ought instead to have added 18 to 15 obtaining 33, then added 21, etc. From some points of view this would be a 'mistake', in spite of the obvious intelligence involved. One can also

imagine a situation where the children were given a number of additions to do, of which the first 5 were all arithmetic progressions, but the 6th was say $23+34+45+\ldots+100+112+122+\ldots+199$. Gauss might have given the answer to this as if it were an arithmetic progression, not having noticed that the 9th term was 112 instead of 111. This would be a definite mistake, which the less intelligent children would not have been likely to make.

The view (d) that intelligence in machinery is merely a reflection of that of its creator is rather similar to the view that the credit for the discoveries of a pupil should be given to his teacher. In such a case the teacher would be pleased with the success of his methods of education, but would not claim the results themselves unless he had actually communicated them to his pupil. He would certainly have envisaged in very broad outline the sort of thing his pupil might be expected to do, but would not expect to foresee any sort of detail. It is already possible to produce machines where this sort of situation arises in a small degree. One can produce 'paper machines' for [[3]] playing chess. Playing against such a machine gives a definite feeling that one is pitting one's wits against something alive.

These views will all be developed more completely below.

VARIETIES OF MACHINERY

It will not be possible to discuss possible means of producing intelligent machinery without introducing a number of technical terms to describe different kinds of existent machinery.

'*Discrete*' *and* '*continuous*' *machinery*. We may call a machine 'discrete' when it is natural to describe its possible states as a discrete set, the motion of the machine occurring by jumping from one state to another. The states of 'continuous' machinery on the other hand form a continuous manifold, and the behaviour of the machine is described by a curve on this manifold. All machinery can be regarded as continuous, but when it is possible to regard it as discrete it is usually best to do so. The states of discrete machinery will be described as 'configurations'.

'*Controlling*' *and* '*active*' *machinery*. Machinery may be described as 'controlling' if it only deals with information. In practice this condition is much the same as saying that the magnitude of the machine's effects may be as small as we please, so long as we do not introduce confusion through Brownian movement, etc. 'Active' machinery is intended to produce some definite physical effect.

Examples	A Bulldozer	Continuous Active
	A Telephone	Continuous Controlling
	A Brunsviga	Discrete Controlling [[4]]
	A Brain (probably)	Continuous Controlling, but is very similar to much discrete machinery

The ENIAC, ACE, etc.	Discrete Controlling
A Differential Analyser	Continuous Controlling.

We shall mainly be concerned with discrete controlling machinery. As we have mentioned, brains very nearly fall into this class, and there seems every reason to believe that they could have been made to fall genuinely into it without any change in their essential properties. However, the property of being 'discrete' is only an advantage for the theoretical investigator, and serves no evolutionary purpose, so we could not expect Nature to assist us by producing truly 'discrete' brains.

Given any discrete machine the first thing we wish to find out about it is the number of states (configurations) it can have. This number may be infinite (but enumerable) in which case we say that the machine has infinite memory (or storage) capacity. If the machine has a finite number N of possible states then we say that it has a memory capacity of (or equivalent to) $\log_2 N$ binary digits. According to this definition we have the following table of capacities, very roughly

Brunsviga	90
ENIAC without cards and with fixed programme	600
ACE as proposed	60,000
Manchester machine (as actually working 8 August 1947)	1,100

The memory capacity of a machine more than anything else determines the complexity of its possible behaviour.

The behaviour of a discrete machine is completely described when we are given the state (configuration) of the machine as a function of the immediately preceding state and the relevant external data.

Logical computing machines (LCMs)

In Turing (1937) a certain type of discrete machine was described. It had an infinite memory capacity obtained in the form of an infinite tape marked out into squares on each of which a symbol could be printed. At any moment there is one symbol in the machine; it is called the scanned symbol. The machine can alter the scanned symbol and its behaviour is in part described by that symbol, but the symbols on the tape elsewhere do not affect the behaviour of the machine. However the tape can be moved back and forth through the machine, this being one of the elementary operations of the machine. Any symbol on the tape may therefore eventually have an innings.

These machines will here be called 'Logical Computing Machines'. They are chiefly of interest when we wish to consider what a machine could in principle be designed to do, when we are willing to allow it both unlimited time and unlimited storage capacity.

Universal logical computing machines. It is possible to describe LCMs in a very standard way, and to put the description into a form which can be

'understood' (i.e., applied by) a special machine. In particular it is possible to design a 'universal machine' which is an LCM such that if the standard description of some other LCM is imposed on the otherwise blank tape from outside, and the (universal) machine then set going it will carry out the operations of the particular machine whose description it was given. For details the reader must refer to Turing (1937).

The importance of the universal machine is clear. We do not need to have an infinity of different machines doing different jobs. A single one will suffice. The engineering problem of producing various machines for various jobs is replaced by the office work of 'programming' the universal machine to do these jobs.

It is found in practice that LCMs can do anything that could be described as 'rule of thumb' or 'purely mechanical'. This is sufficiently well established [[5]] that it is now agreed amongst logicians that 'calculable by means of an LCM' is the correct accurate rendering of such phrases. There are several mathematically equivalent but superficially very different renderings.

Practical computing machines (PCMs)

Although the operations which can be performed by LCMs include every rule-of-thumb process, the number of steps involved tends to be enormous. This is mainly due to the arrangement of the memory along the tape. Two facts which need to be used together may be stored very far apart on the tape. There is also rather little encouragement, when dealing with these machines, to condense the stored expressions at all. For instance the number of symbols required in order to express a number in Arabic form (e.g., 149056) cannot be given any definite bound, any more than if the numbers are expressed in the 'simplified Roman' form (IIIII...I, with 149056 occurrences of I). As the simplified Roman system obeys very much simpler laws one uses it instead of the Arabic system.

In practice however one *can* assign finite bounds to the numbers that one will deal with. For instance we can assign a bound to the number of steps that we will admit in a calculation performed with a real machine in the following sort of way. Suppose that the storage system depends on charging condensers of capacity $C=1\,\mu$f, and that we use two states of charging, $E=100$ volts and $-E=-100$ volts. When we wish to use the information carried by the condenser we have to observe its voltage. Owing to thermal agitation the voltage observed will always be slightly wrong, and the probability of an error between V and $V-\mathrm{d}V$ volts is

$$\frac{2kT}{\pi C}\, e^{-\frac{1}{2}V^2C/kT}\, V\mathrm{d}V$$

where k is Boltzmann's constant. Taking the values suggested we find that the probability of reading the sign of the voltage wrong is about $10^{-1\cdot2\times10^{16}}$. If then a job took more than $10^{10^{17}}$ steps we should be virtually certain of

getting the wrong answer, and we may therefore restrict ourselves to jobs with fewer steps. Even a bound of this order might have useful simplifying effects. More practical bounds are obtained by assuming that a light wave must travel at least 1 cm between steps (this would only be false with a very small machine), and that we could not wait more than 100 years for an answer. This would give a limit of 10^{20} steps. The storage capacity will probably have a rather similar bound, so that we could use sequences of 20 decimal digits for describing the position in which a given piece of data was to be found, and this would be a really valuable possibility.

Machines of the type generally known as 'Automatic Digital Computing Machines' often make great use of this possibility. They also usually put a great deal of their stored information in a form very different from the tape form. By means of a system rather reminiscent of a telephone exchange it is made possible to obtain a piece of information almost immediately by 'dialling' the position of this information in the store. The delay may be only a few microseconds with some systems. Such machines will be described as 'Practical Computing Machines'.

Universal practical computing machines. Nearly all of the PCMS now under construction have the essential properties of the 'Universal Logical Computing Machines' mentioned earlier. In practice, given any job which could have been done on an LCM one can also do it on one of these digital computers. I do not mean that we can do any required job of the type mentioned on it by suitable programming. The programming is pure paper work. It naturally occurs to one to ask whether, e.g., the ACE would be truly universal if its memory capacity were infinitely extended. I have investigated this question, and the answer appears to be as follows, though I have not proved any formal mathematical theorem about it. As has been explained, the ACE at present uses finite sequences of digits to describe positions in its memory: they are actually sequences of 9 binary digits (September 1947). The ACE also works largely for other purposes with sequences of 32 binary digits. If the memory were extended, e.g., to 1000 times its present capacity, it would be natural to arrange the memory in blocks of nearly the maximum capacity which can be handled with the 9 digits, and from time to time to switch
[[6]] from block to block. A relatively small part would never be switched. This would contain some of the more fundamental instruction tables and those concerned with switching. This part might be called the 'central part'. One would then need to have a number which described which block was in action at any moment. However this number might be as large as one pleased. Eventually the point would be reached where it could not be stored in a word (32 digits), or even in the central part. One would then have to set aside a block for storing the number, or even a sequence of blocks, say blocks 1, 2, . . . n. We should then have to store n, and in theory it would be of indefinite size. This sort of process can be extended in all sorts of ways, but we shall always be left with a positive integer which is of indefinite size

and which needs to be stored somewhere, and there seems to be no way out of the difficulty but to introduce a 'tape'. But once this has been done, and since we are only trying to prove a theoretical result, one might as well, whilst proving the theorem, ignore all the other forms of storage. One will in fact have a ULCM with some complications. This in effect means that one will not be able to prove any result of the required kind which gives any intellectual satisfaction.

Paper machines

It is possible to produce the effect of a computing machine by writing down a set of rules of procedure and asking a man to carry them out. Such a combination of a man with written instructions will be called a 'Paper Machine'. A man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal machine. The expression 'paper machine' will often be used below.

Partially random and apparently partially random machines

It is possible to modify the above described types of discrete machines by allowing several alternative operations to be applied at some points, the alternatives to be chosen by a random process. Such a machine will be described as 'partially random'. If we wish to say definitely that a machine is not of this kind we will describe it as 'determined'. Sometimes a machine may be strictly speaking determined but appear superficially as if it were partially random. This would occur if for instance the digits of the number π were used to determine the choices of a partially random machine, where previously a dice thrower or electronic equivalent had been used. These machines are known as apparently partially random.

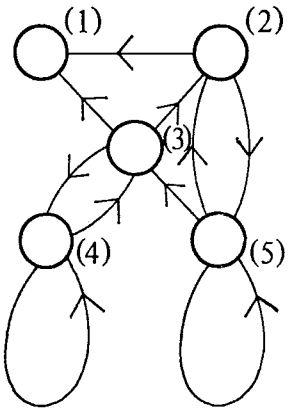
UNORGANIZED MACHINES

So far we have been considering machines which are designed for a definite purpose (though the universal machines are in a sense an exception). We might instead consider what happens when we make up a machine in a comparatively unsystematic way from some kind of standard components. We could consider some particular machine of this nature and find out what sort of things it is likely to do. Machines which are largely random in their construction in this way will be called 'Unorganized Machines'. This does not pretend to be an accurate term. It is conceivable that the same machine might be regarded by one man as organized and by another as unorganized.

A typical example of an unorganized machine would be as follows. The machine is made up from a rather large number N of similar units. Each unit has two input terminals, and has an output terminal which can be connected to the input terminals of (0 or more) other units. We may imagine that for each integer r, $1 \leqslant r \leqslant N$ two numbers $i(r)$ and $j(r)$ are chosen at random

from 1 . . . N and that we connect the inputs of unit r to the outputs of units (r) and $j(r)$. All of the units are connected to a central synchronizing unit from which synchronizing pulses are emitted at more or less equal intervals of time. The times when these pulses arrive will be called 'moments'. Each unit is capable of having two states at each moment. These states may be called 0 and 1. The state is determined by the rule that the states of the units from which the input leads come are to be taken at the previous moment, multiplied together and the result subtracted from 1. An unorganized machine of this character is shown in the diagram below.

r	i(r)	j(r)
1	3	2
2	3	5
3	4	5
4	3	4
5	2	5



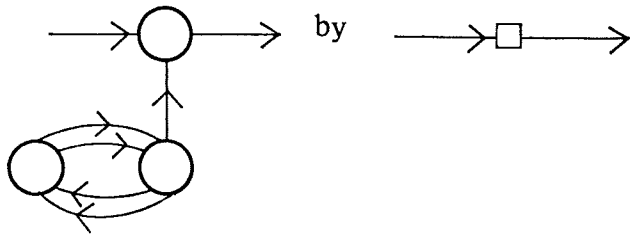
A sequence of six possible consecutive conditions for the whole machine is:

1	1	1	0	0	1	0
2	1	1	1	0	1	0
[7] 3	0	1	1	1	1	1
4	0	1	0	1	0	1
5	1	0	1	0	1	0

The behaviour of a machine with so few units is naturally very trivial. However, machines of this character can behave in a very complicated manner when the number of units is large. We may call these A-type unorganized machines. Thus the machine in the diagram is an A-type unorganized machine of 5 units. The motion of an A-type machine with N units is of course eventually periodic, as in any determined machine with finite memory capacity. The period cannot exceed 2^N moments, nor can the length of time before the periodic motion begins. In the example above the period is 2 moments and there are 3 moments before the periodic motion begins. 2^N is 32.

The A-type unorganized machines are of interest as being about the simplest model of a nervous system with a random arrangement of neurons. It would therefore be of very great interest to find out something about their behaviour. A second type of unorganized machine will now be described, not because it is

of any great intrinsic importance, but because it will be useful later for illustrative purposes. Let us denote the circuit



as an abbreviation. Then for each A-type unorganized machine we can construct another machine by replacing each connection ——→—— in it by ——→□——→. The resulting machines will be called B-type unorganized machines. It may be said that the B-type machines are all A-type. To this I would reply that the above definitions if correctly (but drily!) set out would take the form of describing the probability of an A- (or B-) type machine belonging to a given set; it is not merely a definition of which are the A-type machines and which are the B-type machines. If one chooses an A-type machine, with a given number of units, at random, it will be extremely unlikely that one will get a B-type machine.

It is easily seen that the connection ——→□——→ can have three conditions. It may (i) pass all signals through with interchange of 0 and 1, or (ii) it may convert all signals into 1, or again (iii) it may act as in (i) and (ii) in alternate moments. (Alternative (iii) has two sub-cases.) Which of these cases applies depends on the initial conditions. There is a delay of two moments in going through ——→□——→.

INTERFERENCE WITH MACHINERY. MODIFIABLE AND SELF-MODIFYING MACHINERY

The types of machine that we have considered so far are mainly ones that are allowed to continue in their own way for indefinite periods without interference from outside. The universal machines were an exception to this, in that from time to time one might change the description of the machine which is being imitated. We shall now consider machines in which such interference is the rule rather than the exception.

We may distinguish two kinds of interference. There is the extreme form in which parts of the machine are removed and replaced by others. This may be described as 'screwdriver interference'. At the other end of the scale is 'paper interference', which consists in the mere communication of information to the machine, which alters its behaviour. In view of the properties of the universal machine we do not need to consider the difference between these

two kinds of machine as being so very radical after all. Paper interference when applied to the universal machine can be as useful as screwdriver interference.

We shall mainly be interested in paper interference. Since screwdriver interference can produce a completely new machine without difficulty there is rather little to be said about it. In future 'interference' will normally mean 'paper interference'.

When it is possible to alter the behaviour of a machine very radically we may speak of the machine as being 'modifiable'. This is a relative term. One machine may be spoken of as being more modifiable than another.

One may also sometimes speak of a machine modifying itself, or of a machine changing its own instructions. This is really a nonsensical form of phraseology, but is convenient. Of course, according to our conventions the 'machine' is completely described by the relation between its possible configurations at consecutive moments. It is an abstraction which, by the form of its definition, cannot change in time. If we consider the machine as starting in a particular configuration, however, we may be tempted to ignore those configurations which cannot be reached without interference from it. If we do this we should get a 'successor relation' for the configurations with different properties from the original one and so a different 'machine'.

If we now consider interference, we should say that each time interference occurs the machine is probably changed. It is in this sense that interference 'modifies' a machine. The sense in which a machine can modify itself is even more remote. We may, if we wish, divide the operations of the machine into two classes, normal and self-modifying operations. So long as only normal operations are performed we regard the machine as unaltered. Clearly the idea of 'self-modification' will not be of much interest except where the division of operations into the two classes is made very carefully. The sort of case I have in mind is a computing machine like the ACE where large parts of the storage are normally occupied in holding instruction tables. (Instruction tables are the equivalent in UPCMS of descriptions of machines in ULCMS). Whenever the content of this storage was altered by the internal operations of the machine, one would naturally speak of the machine 'modifying itself'.

MAN AS A MACHINE

A great positive reason for believing in the possibility of making thinking machinery is the fact that it is possible to make machinery to imitate any small part of a man. That the microphone does this for the ear, and the television camera for the eye are commonplaces. One can also produce remote-controlled robots whose limbs balance the body with the aid of servo-mechanisms. Here we are chiefly interested in the nervous system. We could produce fairly accurate electrical models to copy the behaviour of nerves, but there seems very little point in doing so. It would be rather like

putting a lot of work into cars which walked on legs instead of continuing to use wheels. The electrical circuits which are used in electronic computing machinery seem to have the essential properties of nerves. They are able to transmit information from place to place, and also to store it. Certainly the nerve has many advantages. It is extremely compact, does not wear out (probably for hundreds of years if kept in a suitable medium!) and has a very low energy consumption. Against these advantages the electronic circuits have only one counter-attraction, that of speed. This advantage is, however, on such a scale that it may possibly outweigh the advantages of the nerve.

One way of setting about our task of building a 'thinking machine' would be to take a man as a whole and to try to replace all the parts of him by machinery. He would include television cameras, microphones, loudspeakers, wheels and 'handling servo-mechanisms' as well as some sort of 'electronic brain'. This would be a tremendous undertaking of course. The object, if produced by present techniques, would be of immense size, even if the 'brain' part were stationary and controlled the body from a distance. In order that the machine should have a chance of finding things out for itself it should be allowed to roam the countryside, and the danger to the ordinary citizen would be serious. Moreover even when the facilities mentioned above were provided, the creature would still have no contact with food, sex, sport and many other things of interest to the human being. Thus although this method is probably the 'sure' way of producing a thinking machine it seems to be altogether too slow and impracticable.

Instead we propose to try and see what can be done with a 'brain' which is more or less without a body providing, at most, organs of sight, speech, and hearing. We are then faced with the problem of finding suitable branches of thought for the machine to exercise its powers in. The following fields appear to me to have advantages:

(i) Various games, e.g., chess, noughts and crosses, bridge, poker

(ii) The learning of languages

(iii) Translation of languages

(iv) Cryptography

(v) Mathematics.

Of these (i), (iv), and to a lesser extent (iii) and (v) are good in that they require little contact with the outside world. For instance in order that the machine should be able to play chess its only organs need be 'eyes' capable of distinguishing the various positions on a specially made board, and means for announcing its own moves. Mathematics should preferably be restricted to branches where diagrams are not much used. Of the above possible fields the learning of languages would be the most impressive, since it is the most human of these activities. This field seems however to depend rather too much on sense organs and locomotion to be feasible.

The field of cryptography will perhaps be the most rewarding. There is a remarkably close parallel between the problems of the physicist and those of the cryptographer. The system on which a message is enciphered corresponds to the laws of the universe, the intercepted messages to the evidence available, the keys for a day or a message to important constants which have to be determined. The correspondence is very close, but the subject matter of cryptography is very easily dealt with by discrete machinery, physics not so easily.

EDUCATION OF MACHINERY

Although we have abandoned the plan to make a 'whole man', we should be wise to sometimes compare the circumstances of our machine with those of a man. It would be quite unfair to expect a machine straight from the factory to compete on equal terms with a university graduate. The graduate has had contact with human beings for twenty years or more. This contact has been modifying his behaviour pattern throughout that period. His teachers have been intentionally trying to modify it. At the end of the period a large number
[[8]] of standard routines will have been superimposed on the original pattern of his brain. These routines will be known to the community as a whole. He is then in a position to try out new combinations of these routines, to make slight variations on them, and to apply them in new ways.

We may say then that in so far as a man is a machine he is one that is subject to very much interference. In fact interference will be the rule rather than the exception. He is in frequent communication with other men, and is continually receiving visual and other stimuli which themselves constitute a form of interference. It will only be when the man is 'concentrating' with a view to eliminating these stimuli or 'distractions' that he approximates a machine without interference.

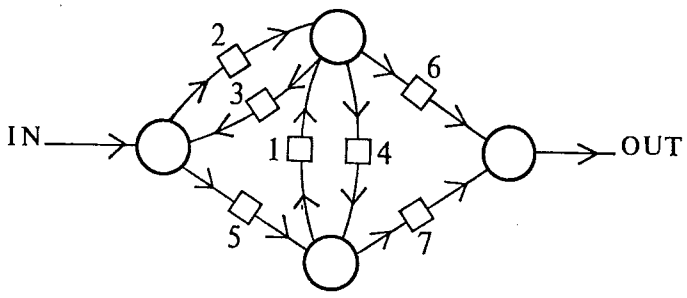
We are chiefly interested in machines with comparatively little interference, for reasons given in the last section, but it is important to remember that although a man when concentrating may behave like a machine without interference, his behaviour when concentrating is largely determined by the way he has been conditioned by previous interference.

If we are trying to produce an intelligent machine, and are following the human model as closely as we can, we should begin with a machine with very little capacity to carry out elaborate operations or to react in a disciplined manner to orders (taking the form of interference). Then by applying appropriate interference, mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands. This would be the beginning of the process. I will not attempt to follow it further now.

ORGANIZING UNORGANIZED MACHINERY

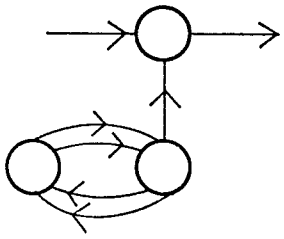
Many unorganized machines have configurations such that if once that configuration is reached, and if the interference thereafter is appropriately

restricted, the machine behaves as one organized for some definite purpose. For instance, the B-type machine shown below was chosen at random.

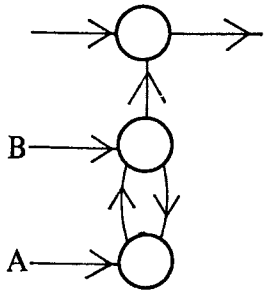


If the connections numbered 1, 3, 6, 4, are in condition (ii) initially and connections 2, 5, 7 are in condition (i), then the machine may be considered to be one for the purpose of passing on signals with a delay of 4 moments. This is a particular case of a very general property of B-type machines (and many other types), viz., that with suitable initial conditions they will do any required job, given sufficient time and provided the number of units is sufficient. In particular with a B-type unorganized machine with sufficient units one can find initial conditions which will make it into a universal machine with a given storage capacity. (A formal proof to this effect might be of some interest, or even a demonstration of it starting with a particular unorganized B-type machine, but I am not giving it as it lies rather too far outside the main argument.)

With these B-type machines the possibility of interference which could set in appropriate initial conditions has not been arranged for. It is however not difficult to think of appropriate methods by which this could be done. For instance instead of the connection



one might use



Here *A*, *B* are interfering inputs, normally giving the signal '1'. By supplying appropriate other signals at *A*, *B* we can get the connection into condition (i) or (ii), as desired. However this requires two special interfering inputs for each connection.

We shall be interested mainly in cases where there are only quite few independent inputs altogether, so that all the interference which sets up the 'initial conditions' of the machine has to be provided through one or two inputs. The process of setting up these initial conditions so that the machine will carry out some particular useful task may be called 'organizing the machine'. 'Organizing' is thus a form of 'modification'.

THE CORTEX AS AN UNORGANIZED MACHINE

Many parts of a man's brain are definite nerve circuits required for quite definite purposes. Examples of these are the 'centres' which control respiration, sneezing, following moving objects with the eyes, etc.: all the reflexes proper (not 'conditioned') are due to the activities of these definite structures in the brain. Likewise the apparatus for the more elementary analysis of shapes and sounds probably comes into this category. But the more intellectual activities of the brain are too varied to be managed on this basis. The difference between the languages spoken on the two sides of the Channel is not due to difference in development of the French-speaking and English-speaking parts of the brain. It is due to the linguistic parts having been subjected to different training. We believe then that there are large parts of the brain, chiefly in the cortex, whose function is largely indeterminate. In the infant these parts do not have much effect: the effect they have is uncoordinated. In the adult they have great and purposive effect: the form of this effect depends on the training in childhood. A large remnant of the random behaviour of infancy remains in the adult.

All of this suggests that the cortex of the infant is an unorganized machine, which can be organized by suitable interfering training. The organizing might result in the modification of the machine into a universal machine or something like it. This would mean that the adult will obey orders given in appropriate language, even if they were very complicated; he would have no common sense, and would obey the most ridiculous orders unflinchingly. When all his orders had been fulfilled he would sink into a comatose state or perhaps obey some standing order, such as eating. Creatures not unlike this can really be found, but most people behave quite differently under many circumstance. However the resemblance to a universal machine is still very great, and suggests to us that the step from the unorganized infant to a universal machine is one which should be understood. When this has been mastered we shall be in a far better position to consider how the organizing process might have been modified to produce a more normal type of mind.

This picture of the cortex as an unorganized machine is very satisfactory

from the point of view of evolution and genetics. It clearly would not require any very complex system of genes to produce something like the A- or B-type unorganized machine. In fact this should be much easier than the production of such things as the respiratory centre. This might suggest that intelligent races could be produced comparatively easily. I think this is wrong because the possession of a human cortex (say) would be virtually useless if no attempt was made to organize it. Thus if a wolf by a mutation acquired a human cortex there is little reason to believe that he would have any selective advantage. If however the mutation occurred in a milieu where speech had developed (parrot-like wolves), and if the mutation by chance had well permeated a small community, then some selective advantage might be felt. It would then be possible to pass information on from generation to generation. However this is all rather speculative.

EXPERIMENTS IN ORGANIZING: PLEASURE–PAIN SYSTEMS

It is interesting to experiment with unorganized machines admitting definite types of interference and try to organize them, e.g., to modify them into universal machines.

The organization of a machine into a universal machine would be most impressive if the arrangements of interference involve very few inputs. The training of the human child depends largely on a system of rewards and punishments, and this suggests that it ought to be possible to carry through the organizing with only two interfering inputs, one for 'pleasure' or 'reward' (R) and the other for 'pain' or punishment' (P). One can devise a large number of such 'pleasure–pain' systems. I will use this term to mean an unorganized machine of the following general character: The configurations of the machine are described by two expressions, which we may call the character-expression and the situation-expression. The character and situation at any moment, together with the input signals, determine the character and situation at the next moment. The character may be subject to some random variation. Pleasure interference has a tendency to fix the character, i.e., towards preventing it changing, whereas pain stimuli tend to disrupt the character, causing features which had become fixed to change, or to become again subject to random variation.

This definition is probably too vague and general to be very helpful. The idea is that when the 'character' changes we like to think of it as a change in the machine, but the 'situation' is merely the configuration of the machine described by the character. It is intended that pain stimuli occur when the machine's behaviour is wrong, pleasure stimuli when it is particularly right. With appropriate stimuli on these lines, judiciously operated by the 'teacher', one may hope that the 'character' will converge towards the one desired, i.e., that wrong behaviour will tend to become rare.

I have investigated a particular type of pleasure–pain system, which I will now describe.

THE P-TYPE UNORGANIZED MACHINE

The P-type machine may be regarded as an LCM without a tape, and whose description is largely incomplete. When a configuration is reached, for which the action is undetermined, a random choice for the missing data is made and the appropriate entry is made in the description, tentatively, and is applied. When a pain stimulus occurs all tentative entries are cancelled, and when a pleasure stimulus occurs they are all made permanent.

Specifically. The situation is a number $s=1, 2, \ldots, N$ and corresponds to the configuration of the incomplete machine. The character is a table of N entries showing the behaviour of the machine in each situation. Each entry has to say something both about the next situation and about what action the machine has to take. The action part may be either

[[9]] (i) To do some externally visible act A_1 or $A_2 \ldots A_K$

(ii) To set one of the memory units $M_1 \ldots M_R$ either into the '1' condition or into the '0' condition.

The next situation is always the remainder either of $2s$ or of $2s+1$ on division by N. These may be called alternatives 0 and 1. Which alternative applies may be determined by either

(a) one of the memory units

(b) a sense stimulus

(c) the pleasure–pain arrangements.

In each situation it is determined which of these applies when the machine is made, i.e., interference cannot alter which of the three cases applies. Also in cases (a) and (b) interference can have no effect. In case (c) the entry in the character table may be either U ('uncertain'), or T0 (tentative 0), T1, D0 (definite 0) or D1. When the entry in the character for the current situation is U then the alternative is chosen at random, and the entry in the character is changed to T0 or T1 according as 0 or 1 was chosen. If the character entry was T0 or D0 then the alternative is 0 and if it is T1 or D1 then the alternative is 1. The changes in character include the above mentioned change from U to T0 or T1, and a change of every T to D when a pleasure stimulus occurs, changes of T0 and T1 to U when a pain stimulus occurs.

We may imagine the memory units essentially as 'trigger circuits' or switches. The sense stimuli are means by which the teacher communicates 'unemotionally' to the machine, i.e., otherwise than by pleasure and pain stimuli. There are a finite number S of sense stimulus lines, and each always carries either the signal 0 or 1.

A small P-type machine is described in the table below

[[10]]

1	P	A	
2	P	B	M1=1
3	P	B	
4	S1	A	M1=0
5	M1	C	

In this machine there is only one memory unit M1 and one sense line S1. Its behaviour can be described by giving the successive situations together with the actions of the teacher: the latter consist of the values of S1 and the rewards and punishments. At any moment the 'character' consists of the above table with each 'P' replaced by either U, T0, D0 or D1. In working out the behaviour of the machine it is convenient first of all to make up a sequence of random digits for use when the U cases occur. Underneath these we may write the sequence of situations, and have other rows for the corresponding entries from the character, and for the actions of the teacher. The character and the values stored in the memory units may be kept on another sheet. The T entries may be made in pencil and the D entries in ink. A bit of the behaviour of the machine is given below:

Random sequence	0	0	1	1	1	0	0	1	0	0	1	1	0	1	1	0	0	0
Situations	3	1	3	1	3	1	3	1	2	4	4	4	3	2	.	.		
Alternative given by		U	T	T	T	T	T	U	U	S	S	S	U	T				
			0	0	0	0	0			1	1	1		0				
Visible action	B	A	B	A	B	A	B	A	B	A	A	A	B	B				
Rew. & Pun.							P											
Changes in S1	1											0						

It will be noticed that the machine very soon got into a repetitive cycle. This became externally visible through the repetitive BABAB . . . . By means of a pain stimulus this cycle was broken.

It is probably possible to organize these P-type machines into universal machines, but it is not easy because of the form of memory available. It would be necessary to organize the randomly distributed 'memory units' to provide a systematic form of memory, and this would not be easy. If, however, we supply the P-type machine with a systematic external memory this organizing becomes quite feasible. Such a memory could be provided in the form of a tape, and the externally visible operations could include movement to right and left along the tape, and altering the symbol on the tape to 0 or to 1. The sense lines could include one from the symbol on the tape. Alternatively, if the memory were to be finite, e.g., not more than 2^{32} binary digits, we could use a dialling system. (Dialling systems can also be used with an infinite memory, but this is not of much practical interest.) I have succeeded in organizing such a (paper) machine into a universal machine.

The details of the machine involved were as follows. There was a circular memory consisting of 64 squares of which at any moment one was in the machine ('scanned') and motion to right or left were among the 'visible actions'. Changing the symbol on the square was another 'visible action', and the symbol was connected to one of the sense lines S1. The even-numbered squares also had another function, they controlled the dialling of information to or from the main memory. This main memory consisted of 2^{32} binary

digits. At any moment one of these digits was connected to the sense line S2. The digit of the main memory concerned was that indicated by the 32 even positioned digits of the circular memory. Another two of the 'visible actions' were printing 0 or 1 in this square of the main memory. There were also three ordinary memory units and three sense units S3, S4, S5. Also six other externally visible actions A,B,C,D,E,F.

This P-type machine with external memory has, it must be admitted, considerably more 'organization' than say the A-type unorganized machine. Nevertheless the fact that it can be organized into a universal machine still remains interesting.

The actual technique by which the 'organizing' of the P-type machine was carried through is perhaps a little disappointing. It is not sufficiently analogous to the kind of process by which a child would really be taught. The process actually adopted was first to let the machine run for a long time with continuous application of pain, and with various changes of the sense data S3, S4, S5. Observation of the sequence of externally visible actions for some thousands of moments made it possible to set up a scheme for identifying the situations, i.e., by which one could at any moment find out what the situation was, except that the situations as a whole had been renamed. A similar investigation, with less use of punishment, enables one to find the situations which are affected by the sense lines; the data about the situations involving the memory units can also be found but with more difficulty. At this stage the character has been reconstructed. There are no occurrences of T0, T1, D0, D1. The next stage is to think up some way of replacing the 0s of the character by D0, D1 in such a way as to give the desired modification. This will normally be possible with the suggested number of situations (1000), memory units, etc. The final stage is the conversion of the character into the chosen one. This may be done simply by allowing the machine to wander at random through a sequence of situations, and applying pain stimuli when the wrong choice is made, pleasure stimuli when the right one is made. It is best also to apply pain stimuli when irrelevant choices are made. This is to prevent getting isolated in a ring of irrelevant situations. The machine is now 'ready for use'.

The form of universal machine actually produced in this process was as follows. Each instruction consisted of 128 digits, which we may regard as forming four sets of 32, each of which describes one place in the main memory. These places may be called P,Q,R,S. The meaning of the instruction is that if p is the digit at P and q that at Q then $1-pq$ is to be transferred to position R and that the next instruction will be found in the 128 digits beginning at S. This gives a UPCM, though with rather less facilities than are available say on the ACE.

I feel that more should be done on these lines. I would like to investigate other types of unorganized machines, and also to try out organizing methods that would be more nearly analogous to our 'methods of education'. I made

a start on the latter but found the work altogether too laborious at present. When some electronic machines are in actual operation I hope that they will make this more feasible. It should be easy to make a model of any particular machine that one wishes to work on within such a UPCM instead of having to work with a paper machine as at present. If also one decided on quite definite 'teaching policies' these could also be programmed into the machine. One would then allow the whole system to run for an appreciable period, and then break in as a kind of 'inspector of schools' and see what progress had been made. One might also be able to make some progress with unorganized machines more like the A- and B-types. The work involved with these is altogether too great for pure paper-machine work.

One particular kind of phenomenon I had been hoping to find in connection with the P-type machines. This was the incorporation of old routines into new. One might have 'taught' (i.e., modified or organized) a machine to add (say). Later one might teach it to multiply by small numbers by repeated addition and so arrange matters that the same set of situations which formed the addition routine, as originally taught, was also used in the additions involved in the multiplication. Although I was able to obtain a fairly detailed picture of how this might happen I was not able to do experiments on a sufficient scale for such phenomena to be seen as part of a large context. [[11]]

I also hoped to find something rather similar to the 'irregular verbs' which add variety to language. We seem to be quite content that things should not obey too mathematically regular rules. By long experience we can pick up and apply the most complicated rules without being able to enunciate them at all. I rather suspect that a P-type machine without the systematic memory would behave in a rather similar manner because of the randomly distributed memory units. Clearly this could only be verified by very painstaking work; by the very nature of the problem 'mass production' methods like built-in teaching procedures could not help.

DISCIPLINE AND INITIATIVE

If the untrained infant's mind is to become an intelligent one, it must acquire both discipline and initiative. So far we have been considering only discipline. To convert a brain or machine into a universal machine is the extremest form of discipline. Without something of this kind one cannot set up proper communication. But discipline is certainly not enough in itself to produce intelligence. That which is required in addition we call initiative. This statement will have to serve as a definition. Our task is to discover the nature of this residue as it occurs in man, and to try and copy it in machines.

Two possible methods of setting about this present themselves. On the one hand we have fully disciplined machines immediately available, or in a matter of months or years, in the form of various UPCMS. We might try to graft some initiative onto these. This would probably take the form of programming the machine to do every kind of job that could be done, as a

matter of principle, whether it were economical to do it by machine or not. Bit by bit one would be able to allow the machine to make more and more 'choices' or 'decisions'. One would eventually find it possible to program it so as to make its behaviour be the logical result of a comparatively small number of general principles. When these became sufficiently general, interference would no longer be necessary, and the machine would have 'grown up'. This may be called the 'direct method'.

The other method is to start with an unorganized machine and to try to bring both discipline and initiative into it at once, i.e., instead of trying to organize the machine to become a universal machine, to organize it for initiative as well. Both methods should, I think, be attempted.

Intellectual, genetical and cultural searches

A very typical sort of problem requiring some sort of initiative consists of those of the form 'Find a number n such that . . .'. This form covers a very great variety of problems. For instance problems of the form 'See if you can find a way of calculating the function which will enable us to obtain the values for arguments . . . to accuracy . . . within a time . . . using the UPCM . . .' are reducible to this form, for the problem is clearly equivalent to that of finding a program to put on the machine in question, and it is easy to put the programs into correspondence with the positive integers in such a way that given either the number or the program the other can easily be found. We should not go far wrong for the time being if we assumed that all problems were reducible to this form. It will be time to think again when something turns up which is obviously not of this form.

The crudest way of dealing with such a problem is to take the integers in order and to test each one to see whether it has the required property, and to go on until one is found which has it. Such a method will only be successful in the simplest cases. For instance in the case of problems of the kind mentioned above, where one is really searching for a program, the number required will normally be somewhere between 2^{1000} and $2^{1,000,000}$. For practical work therefore some more expeditious method is necessary. In a number of cases the following method would be successful. Starting with a UPCM we first put a program into it which corresponds to building in a logical system (like Russell's *Principia Mathematica*). This would not determine the behaviour of the machine completely: at various stages more than one choice as to the next step would be possible. We might arrange, however, to take all possible arrangement of choices in order, and go on until the machine proved
[[12]] a theorem, which, by its form, could be verified to give a solution of the problem. This may be seen to be a conversion of the original problem into another of the same form. Instead of searching through values of the original variable n one searches through values of something else. In practice when solving problems of the above kind one will probably apply some very complex 'transformation' of the original problem, involving searching through

various variables, some more analogous to the original one, some more like a 'search through all proofs'. Further research into intelligence of machinery will probably be very greatly concerned with 'searches' of this kind. We may perhaps call such searches 'intellectual searches'. They might very briefly be defined as 'searches carried out by brains for combinations with particular properties'.

It may be of interest to mention two other kinds of search in this connection. There is the genetical or evolutionary search by which a combination of genes is looked for, the criterion being survival value. The remarkable success of this search confirms to some extent the idea that intellectual activity consists mainly of various kinds of search.

The remaining form of search is what I should like to call the 'cultural search'. As I have mentioned, the isolated man does not develop any intellectual power. It is necessary for him to be immersed in an environment of other men, whose techniques he absorbs during the first twenty years of his life. He may then perhaps do a little research of his own and make a very few discoveries which are passed on to other men. From this point of view the search for new techniques must be regarded as carried out by the human community as a whole, rather than by individuals.

INTELLIGENCE AS AN EMOTIONAL CONCEPT

The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence. With the same object therefore it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour.

It is possible to do a little experiment on these lines, even at the present stage of knowledge. It is not difficult to devise a paper machine which will play a not very bad game of chess. Now get three men as subjects for the experiment A,B,C. A and C are to be rather poor chess players, B is the operator who works the paper machine. (In order that he should be able to work it fairly fast it is advisable that he be both mathematician and chess player.) Two rooms are used with some arrangement for communicating moves, and a [[13]] game is played between C and either A or the paper machine. C may find it quite difficult to tell which he is playing. (This is a rather idealized form of an experiment I have actually done.)

REFERENCES

Church, Alonzo (1936) An unsolvable problem of elementary number theory. *Amer. J. of Math.* **58,** 345–63.

Gödel, K. (1931) Über formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme. *Monatshefte fur Math. und Phys.* **38,** 173–89.

Turing, A.M. (1937) On computable numbers with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* **42,** 230–65.