Journal of the British Society for Phenomenology

ISSN: 0007-1773 (Print) 2332-0486 (Online) Journal homepage: www.tandfonline.com/journals/rbsp20

My Reply to Turing: Fiftieth Anniversary

Wolfe Mays

To cite this article: Wolfe Mays (2001) My Reply to Turing: Fiftieth Anniversary, Journal of the British Society for Phenomenology, 32:1, 4-23, DOI: 10.1080/00071773.2001.11007314

To link to this article: https://doi.org/10.1080/00071773.2001.11007314

Journal of the British Society for Phenomenology, Vol. 32, No. 1, January 2001

MY REPLY TO TURING: FIFTIETH ANNIVERSARY

WOLFE MAYS

Introduction

In the summer of 1950 Gilbert Ryle sent me the galley proofs of Aean Turing's article "Computing Machines and Intelligence" which was to appear in the October number of *Mind* 1950.[1] In this paper Turing largely concerned himself with the question "Can a machine think?". This year then marks the 50th anniversary of its appearance, and is being celebrated as such by cognitive scientists and workers in the field of artificial intelligence. Ryle asked me whether I would write a reply. This I did and delivered it in good time and thought I had made a decent job of it. To my surprise he wrote back rejecting my piece on the ground that it was too polemical. However, it got published in *Philosophy* a little later.[2]

Re-reading it I am struck by the fact that though I still think my criticisms were justified, I must have been both philosophically and politically naive to express them in the way I did. Ryle had only a year earlier written a book called *The Concept of Mind*[3] in which he tried to exorcise the ghost in the machine, that is to say, consciousness, thoughts, feelings, desires, etc., all examples of what he called the category mistake or reification. Instead, he tried to define them in terms of our behavioural activities, or "knowing-how". Ryle must have been somewhat annoyed on reading my reply. What I was doing was committing the mortal sin of trying to resurrect the ghost!

John Searle has pointed out that "As recently as a few years ago if one raised the subject of consciousness in cognitive science discussions, it was generally regarded as a form of bad taste, and graduate students who are always attuned to the social mores of their disciplines, would raise their eyes to the ceiling and assume expressions of mild disgust".[4] There is now a greater tolerance to using the word consciousness. We even have a *Journal of Consciousness Studies.* But it would have been difficult for Ryle to admit that explanations of our higher mental processes in terms of consciousness, might give us a better understanding of them than would explanations in behavioural terms. To do so would undermine his whole attack on the ghost in the machine. Nevertheless, Ryle's thought still lives on in the work of Daniel Dennet and others.

Why did Ryle ask me to reply in the first place? I believe he thought I would take a somewhat similar line to that of Turing, although obviously he did not expect me to be in complete agreement with him. I ought to mention that in 1949 I had helped to construct an electrical logic machine and demonstrated it in July 1950 at the *Joint Session of Mind Association and*

Aristotelian Society at Bristol. He may therefore have taken it that I was also interested in mechanising thought activities, especially as at that time Turing was a colleague of mine at Manchester.

Turing's views on mind and thought are somewhat speculative as he admitted in his *Mind* paper. Thus he tells us, "I have no very convincing arguments of a positive nature to support my view. If I had I should not have taken such pains to point out the fallacies in contrary views".[5] These views largely concern arguments of the sort that have, for example, been expressed by Michael Polanyi. According to Polanyi all machine operations involve reference to the mental activities of the person operating the machine and also to those of its designer. As these activities are not specifiable in any finite number of steps, they cannot therefore be simulated by a machine.[6] This argument has been strengthened by an appeal to Gödel's theorem which it is claimed shows that the mind is able to carry out activities which a machine cannot.[7] Turing did not deal with this objection in his paper.

Can Machines Think?

What follows below is a somewhat revised version of my reply. In the fifty years since it was written there have been dramatic advances in the field of computer technology. A vast literature has also been generated on the mind-machine problem. Because of this parts of my reply need updating. Where this is necessary my updating comments will be placed between square brackets. Nevertheless, I think it was a reasonable response to Turing's paper, and that it still has some relevance today.

I. Machines, thinking and definition.

Mr. A.M. Turing was quoted in *The Times* about a year ago as saying it would be interesting to discover the degree of intellectual activity of which a machine was capable and to what extent it could think for itself. He has now pressed this suggestion further and given the results of his researches in an article called "Computing Machines and Intelligence", together with a brief account of a "child-machine" which he has attempted to educate. (*Mind,* October 1950) I intend to discuss this article in some detail and examine his claim that "machines can think".

Apparantly, his machine will have the following attributes:

1. Think, write, play games and remember.
2. Make decisions, suitable and unsuitable.
3. Observe the results of its own behaviour.
4. Achieve a purpose.
5. Learn by rewards and punishments.
6. Obey commands.
7. Deliberately introduce mistakes in its working.

Turing protests that it is absurd and dangerous to suppose that the answer to the question "Can machines think?", is to be sought in a study of the way these words are normally used. In its place he substitutes a definition in terms of the machine's capacity to play a game which seems to be a variation of the radio game of Twenty Questions, the part of the witness being taken by a computing machine, the part of the interrogator by a human being. Machines able to play the game do not as yet exist, but Turing is firmly confident that in fifty years time it will be possible to programme computers to play the imitation game. He therefore replaces the original question, "Can machines think?" by the question, "Are there imaginable digital computers which would do well in the imitation game?" (i.e., deceive the questioner into believing that it was a human being.)

Turing believes linguistic usage will have altered so much by the end of the century that one will be able to speak of machines "thinking" without expecting to be contradicted. [We have now entered a new millenium and people still disagree as to whether machines can think.] However, if we were merely concerned with the definition or use which Turing gives to the expression "machines can think", or the state of the English language in A.D. 2000, there would be no problem at all for us. People would not feel surprised and a little hurt by this suggestion which brings up, to my mind at least, the image of metallic masses cogitating in a Rodin-like manner. The problem arises because we normally associate thinking with certain conscious processes.

What is, however, of some importance, as well as interest, is to know how people actually use the words today and not how they may use them at some future date. As we are unable to embark upon a Gallup or MORI survey, the simplest method is to consult the dictionary; despite obvious shortcomings, it gives a fair indication of the way these words are normally used.

In the case of a machine, the most relevant of the O.E.D. definitions is that it is "a combination of parts moving mechanically as contrasted with a being having life, consciousness and will. Hence applied to a person who acts merely from habit or obedience to a rule, without intelligence, or to one whose actions have the undeviating precision of a machine".[8] [Although a digital computer is a somewhat different machine than, say, a steam-engine, as it works electronically rather than mechanically, the principle is still the same.]

From the point of view of modern neo-behaviourism this definition may seem a little old fashioned. Perhaps it is well to make a confession of faith here. I accept the evidence of my own introspections, as well as those of other people, that there are such things as private psychological events, however heretical such a view may seem today. The O.E.D. definition does bring out one thing at least: a machine is usually thought of as something which does not possess a private life of its own, it does not indulge in day-dreams when at its task, it lacks consciousness, intelligence and will.

In the light of the above, one sees the point of Professor Jefferson's remark, "When we hear it said that wireless valves think, we may despair of language".[9] It is precisely because by a machine we mean something which does not possess intelligence or consciousness that we boggle at the assertion "machines can think", despite Turing's attempt to stream-line it. In a sense it is a puzzle of our own making: machines are defined as not having precisely those characteristics of thought, feeling and will, we assign to a human being. We have expresssly ruled out any internal private life. Though it may duplicate our overt or external behaviour, it is assumed that it cannot duplicate our internal activities.

In other words, if it repeated statements such as "I feel toothache", "I enjoy strawberries and cream","I hate Mr. X", we would not attach the same significance to them as we would if these statements were made by a human being. They symbolize a privacy of experience which we do not normally attribute to machines. We might suppose there was a gramophone record inside.

As John Locke and, for that matter, Descartes pointed out, if we found a parrot who talked and argued like a man, we would be reluctant to admit that it exercised conscious thought (or even that it was capable of manifesting linguistic skills), whereas we would still be inclined to attribute some sort of mental life even to autistic individuals.

We usually reserve the term "thinking" for human beings; who have a peculiar complex of mental characteristics, who show certain patterns of behaviour and who are not only able to think, but sense, feel and will as well, and to think and will because they sense and feel. Discussion of thinking in the abstract obscures the fact that thinking, at least on its psychological side, is a complex activity in which feeling and volition are inextricably intermingled. One of the few places where pure thought is to be found is in Immanuel Kant's *Critique of Pure Reason.* And perhaps not even there.

If, however, we came across an artifice which showed every sign of intelligent behaviour, and was yet in appearance very unlike a human being, we would nevertheless hesitate before asserting it was capable of having psychological experiences. We might not be able to make up our mind whether it was a living thing (an alien from Mars) or some sort of mechanism. Biologists are familiar with such situations in their taxonomies.

It may therefore be necessary to introduce a new term to indicate a device which simulates overt human activities without at the same time duplicating our internal behaviour. The word is ready to hand and was coined by the Czech Karel Capek in his play R.U.R. We call them "robots", those devices which fall in the twilight zone between man and the normal run of machines; devices, which as far as simulating human beings is concerned, are still in the realm of fiction – in the imagination of their authors.

[We now, it is true, use computer controlled devices for performing such skilled activities as machine tool making, and recently a device for performing simple surgical tasks has been invented. Artefacts have been constructed which will do our cleaning and even wait on us at table. Some cute quasi-animals which will run around laboratory mazes have also been produced. But these are still far from the male or female androids pictured in science fiction.]

In this connection it might be a good thing to drop the word "machine", with its affective overtones of clanging metal, and use some such neutral word as "artifice". Machines which can perform mathematical and logical operations are very different from the internal combustion engines, printing presses and looms met with in our everyday excursions.

The paradoxical Frankenstein nature of the machine-mind arises from the intrinsic difference in our conception of minds and machines. At least for most of us, minds have a certain privacy about them. I can directly inspect the contents of my own mind, but not that of my neighbours, which is, no doubt, all to the good.

The difficulty vanishes, of course, on Turing's definition, but then the meaning of the word "thinking" has changed to such an extent that it has little in common with what we normally mean by it. If a machine could perform this or that human function it would not be what we now mean by a machine. Its meaning has been stretched to such an extent that we might even seriously contemplate calling it a new type of organism. An example of this is seen at the end of R.U.R. where a pair of robots (male and female) develop inner lives of their own. They find themselves at night alone on a hill top. As the dawn slowly breaks their fingers meet – a thrill runs through them – they have fallen in love. Although we deal here with the realm of dramatic fiction, the imaginative cybernetician is indulging in a similar practice, when he attributes thought and emotion to his hypothetical devices.

Even if it were possible to construct a machine whose behaviour was indistinguishable from that of a human being, and even if we accept the behaviourist criterion it might still be useful to distinguish between men (and women) produced by natural methods and artificial men (and women). Were this to occur our social judgments might alter, the same shame might be attached to being born the natural way as we nowadays attach to illegitimacy. A different label might indeed become a social necessity. [The above example does not now seem to work, as over the last 50 years this distinction seems to have become eroded.]

[This was also written before genetic engineering appeared on the horizon. It is probably in this field rather than on the computing machine one, that such possibilities may be realized. Stephen Hawkins (the cosmologist) no doubt had this in mind when he said recently that "the human race needs to increase its complexity if biological systems are to keep ahead of electronic ones. At the moment computers have the advantage of speed, but they show no sign of intelligence. This is not suprising since our present computers are less complex than the brain of an earthworm, a species not known for its intellectual powers".[10]]

II. Consciousness, feeling and machines

In the true Cartesian manner nearly half of Turing's paper (twelve out of the twenty-seven pages) consists in answering objections. One of the objections considered is the "argument from consciousness". As an instance of this he quotes from Professor Jefferson's Lister Oration of 1949.

I give the relevant quotation from Jefferson below:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by a chance fall of symbols, could we agree that machine equals brain. (445)

Turing blandly states that such a view logically leads to solipsism and gives Jefferson the alternatives of either accepting solipsism or his own definition of thinking in terms of the imitation game. There is, however, no connection between the solipsist position and what he calls the argument from consciousness. Though this argument asserts that we are directly aware of our own internal states, it does not exclude indirect knowledge of other people's minds, and it does not even have to assume that we have certain knowledge of our own minds. The alternatives are not exclusive and Jefferson need not accept either. There is also such a thing as empathy where one seems to become directly aware of other people's feelings.

Further, if I understand Jefferson rightly, he is not saying that machines will not be able to write poetry, but that he does not feel justified in describing their performance as thinking until they write poetry or compose concertos, or for that matter construct mathematical theorems, because of thoughts and emotions felt; *felt* is the operative word here. Turing apparantly takes this as an argument for solipsism. It seems rather to be a causal statement describing how creative minds produce poetry, music and mathematics. The type of persons who bear the strongest resemblance to a digital computer, are the so-called lightning calculators, who are often uneducated men, and even sometimes feeble-minded.

From what has already been said, it will be seen that the question "Can machines think?" means something very different for Turing than it does for Professor Jefferson. For Jefferson, and I should say for most ordinary people, any definition of the word "thinking" would also include certain psychological characteristics. Turing and Jefferson are in fact speaking two different languages; in the behaviouristic (or physical) language of Turing – words which only have an objective physical content appear (or should appear), electronic tubes, flip-flop circuits, programmes, etc. [This needs updating in terms of pentiums, hard discs and the appropriate software] It is a deterministic machine language in the grand manner of nineteenth-century Newtonian physics, although this is somewhat obscured by the everyday phraseology in terms of which Turing expresses his position. In the psychological language used by Jefferson words like consciousness, free-will, decision, etc., occur, but words like desiring and feeling ought by their very nature to be prohibited from Turing's account.

When we describe the functioning of mechanisms we need to remain within the bounds of the physical language, and not include in our descriptions subjective words and phrases from the psychological one, otherwise we shall find our account substantially infected by anthropomorphism. Turing, in his account of the wonders of digital computers does this at every stage and turn. He talks of "machines making decisions", of "being punished and rewarded", "deliberately introducing mistakes", "doing homework", "obeying orders", etc. And he concludes his article with the following valediction:

> It can also be maintained that it is best to provide the machine with the best sense-organs that money can buy, and then teach it to understand and speak English". He goes on "This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried. (460)

It would indeed be idle for me to count the number of words Turing uses from the psychological language in his article; none of which he has attempted to redefine objectively as he has with thinking. They are still heavily loaded with emotive and subjective content, and have an essential reference to private psychological states. When he talks of machines being punished and rewarded no doubt he wishes to compare their re-programming with the way young children are taught to acquire new skills. But one ought only talk of rewarding and punishing a machine if like the child it was able to appreciate that its actions may be followed by pleasant or unpleasant consequences.

Any attempt to describe the behaviour of machines (defined in terms of overt behaviour) by means of such a subjective vocabulary will make confusion worse confounded. It is for this reason that it does not make sense to talk of machines having the following psychological characteristics. They are the disabilities of machines mentioned on page 447, and occur in the argument he criticises, which takes the form: "I grant you that you can make machines do all the things you have mentioned but you will never succeed to make one to do *X*".

It will be observed that X is generally an item from the psychological language or a value judgment. Turing gives the following examples: "Be kind, resourceful, beautiful, friendly..., have initiative, have a sense of humour, tell right from wrong, make mistakes..., fall in love, enjoy strawberries and cream...."

Although it might be possible to make a machine enjoy strawberries and cream, any attempt to do so, he thinks, would be idiotic. No reasons are given for this assertion. Useless perhaps, but, if it were possible, would it not be another marvel in the thesaurus of the mechanical necromancer? Or does Turing implicitly recognize that it would be meaningless to talk of machines "enjoying anything," since the word "enjoyment" has a meaning only by reference to our private feelings, no place for which, however, can be found in the physical language, from which they have been expressly excluded since Galileo's time. As Hartree (another computer expert) warns us, the "specialized use of words already current may lead to misunderstanding, particularly when words habitually used in connection with living organisms, and especially with human activities, such as "memory", "choice", "judgment", are applied to mechanism".[11]

III. The limits of machines and logical inference

When Turing wants to know whether there are imaginable computers which could do well on the imitation game he asks the following question:

> Is it true that by modifying the computer to have an adequate storage, suitably increasing its speed of action and providing it with an adequate programme, *C* (the computer) can be made to play satisfactorily the part of *A* (the feigned human being) in the imitation game, the part of *B* being played by a man? (442)

Now this is a question to which the answer true or false cannot be given (at least until fifty years time, i.e., in the year 2000). It asserts a proposition about a hypothetical state of affairs. All that can be said is that we do not just know, as this is a truth depending upon the physical actualization and not upon the conceptual possibility of this state of affairs. As Kant pointed out a long time ago we cannnot argue from a logical possibility to the possibility of the existence of a real thing.

[Fifty years have now gone by and we still do not seem to have succeeded in constructing such a machine to perform the imitation game postulated by Turing, possibly because no one thought it worth trying. I recognise that we can now programme computers to play a good game of chess, and sometimes beat world-masters. Indeed the fact that computers do not seem to have feelings may give them an advantage over a human opponent.]

In principle there is perhaps no reason why an elephant should not have wings and fly, provided they were large enough to give it air-lift; there is no logical reason why we should not all live to be as old as Methuselah, or for survival after death, but there may be important physical limitations, spatial, temporal and mechanical, why this is not possible. I believe, but I may be mistaken in this, that engineers are already finding that there is an optimum limit in the size of computing machines. If this should prove to be the case the computing machine (which is going to imitate the brain) as large as the Empire State Building and powered by the Niagara Falls, will still remain a subject for conjecture.

[I recognise that we now have the silicon chip which has led to a dramatic miniturisation and increase in the power of computers, but we are still a long way from achieving a computer which will precisely imitate the working of the brain. As Hawkins has noted, the nervous system of an earthworm is already more complex than the present state of the art computer. Perhaps brains are the best imitators of themselves.]

Lewis Mumford points out that certain machines have already reached the limit of their development, for example, the printing press, the water turbine and even the telephone system, the only gain is one of cheapness and universality ... "there are bounds to mechanical progress within the nature of the physical world itself. It is only by ignoring the limiting conditions that a belief in the automatic and inevitable and limitless extension of the machine can be maintained".[12]

[Mumford may have underestimated the possibilities of technological advance, but in principle he seems to be correct. The introduction of electronics, circuit miniturisation and improvements in computer software have pushed these limits further back. A good example of this is to be seen in tele-communications, where a fibre optic cable can carry many more messages than a copper wire one could. There is also the internet and with it the world wide web which are now changing our very life-style. But this does not mean that there are no limits to mechanisation and scientific advance. Bertrand Russell, who took up a similar stance to Mumford's, dubbed the belief that one can obtain unlimited control over nature "cosmic impiety".[13] It is also still true that if you want a good suit of clothes or shoes made, they need to be made by hand, and despite computerised type-setting, a good page of print probably would still need to be set up in this way.]

Turing tells his readers that digital computers "can in fact mimic the actions of a human computer very closely" (438). His statement of fact seems to be somewhat dubious, unless he is using mimicry in a very peculiar sort of way where it can have no reference to "intentional imitation". As a large part of his case rests on his conception of "mimicry" it is necessary to give an analysis of it.

For Turing it consists of three stages:

(1) Asking a human computer how he performs the arithmetical operations of addition, subtraction, multiplication, etc.

(2) Coding the answer in the form of an instruction table in terms of a string of 1's and 0's.

(3) Inserting the programme in the machine.

"Mimicking" on his account is then a complex operation and consists in a human programmer translating his answers into the binary language, and then inserting the coded programme into the machine.

Conceivably, (1), (2), and (3) could be done by another machine, but then we are faced with a regress until we come to a machine where the programme was made and inserted by a human being. Indeed one could in the same way talk of a pianola (or a similar instrument) mimicking the behaviour of a pianist, the analogy is in fact precise.

The phrase "digital computer" seems to be used in a rather ambiguous way. Sometimes it refers to the machine standing on its feet, as it were (i.e., the hardware), and sometimes to the loaded machine with a programme (or software) inserted into it. These are two very different things. Even if we take the latter interpretation, and even if we erase from our memories that it has been programmed by a human agent, it still could not be said to be mimicking a human being, as mimicking means intentional imitation. If Turing wishes to restrict himself to a description of the observed behaviour of the machine, without entering into any discussion of its causation, the most he is entitled to claim is that there is a relation of similarity between its behaviour and human behaviour, and not that the relationship is one of "mimicking". "Mime" in its original use as seen, for example, in ballet, was rather an attempt to communicate affective states by overt symbolism.

Mimicking as used by Turing might be regarded as a form of translation, as he talks of the information being coded into a string of 1's and O's, namely, in terms of Boolean algebra, which has a simple vocabulary and grammatical structure. However, there are limitations to the art of translation, which apply to simple logical languages as well as to the more complex natural ones. Natural languages, owing to differences in physical structure and vocabulary are not precisely translatable into each other – "traduttore traditore". It has, for example, been pointed out that the German word order can alter the expressive meaning of even such a simple sentence as " I have always loved you". There is a difference between "Ich habe dich immer geliebt", "Dich habe ich immer geliebt", "Immer habe ich dich geliebt" and "Geliebt habe ich dich immer", which cannot be replicated in English.

Turing suggests that instead of trying to imitate an "adult mind" it might be a good thing to produce a programme which simulates the child's mind. By subjecting it to an appropriate course of instruction we would arrive at the "adult brain." The essentials of this projected course are sketched on pages 456-59 and closely resemble the treatment meted out to a

"Bokanovsky Group" of infants in the Neo-Pavlovian Conditioning Room in Chapter II of Aldous Huxley's *Brave New World.*

"Adult mind" and "adult brain" are used by Turing interchangeably (cf., 456, lines 2 and 4), there is a tacit assumption of identity. Presumably, he tells us, "the child-brain is something like a note-book, as one buys it from the stationers. Rather little in the way of mechanism, and lots of blank sheets". He hopes that the child-brain will turn out to have little in the way of mechanism so that something like it can be easily programmed. [This was written before Chomsky put forward the view that there was such a thing as a deep grammar which was an innate feature of our minds.]

On the face of it Turing's account bears a strong resemblance to the "tabula rasa theory" of John Locke, that the mind at birth is like a wax tablet, its characteristics being impressed upon it by environment and education. Locke's "tabula rasa theory" together with the principle of association was indeed an attempt to apply Newton's physical methods to the realm of mind. Compared with the more Platonic view where education is conceived as a drawing out of the child's potentialities rather than the injection of information, the "tabula rasa" theory shows up badly – it appears inadequate and gives an oversimplified picture. The child's mind may have a good deal more internal structure than Turing bargained for.

[A cognitive scientist could argue that the only difference between these two forms of instruction is that in one the programming is largely dependent on external influences, and in the other it is implicit in the child's genetic structure. But despite the attempt to chart the human genome, we have not yet produced a genetic map of our intellectual abilities.]

What on Turing's view corresponds to the human mind is, however, not just the machine, but the machine plus the instructions (i.e., software) fed into it. Without the programme it cannot be compared to a mind at all. With the programme it is no longer a "tabula rasa". An uncharitable critic might say that it now contains a selected group of innate ideas inserted not by a benevolent Deity, but by a human programmer.

In physics mechanical models with their deterministic structure and misleading pictorial suggestions have fallen into disrepute and been replaced by abstract probability and statistical concepts. The superseding of the Newtonian machine model has led to enormous advances in physics, and yet, curiously enough, just when this happened, its ghost has appeared to haunt the councils of biologists and psychologists.

Machine analogies are in fact a variety of animism. Professor Ryle points out that there are very few natural machines in nature; "inventing machines is not copying things found in inanimate Nature".[14] If we want to find examples in nature of "self-maintaining routine observing systems", we have rather to look to living organisms. The machine then is not something superior to the men and women who construct it, it is in fact an inferior type of animal, "a sort of minor organism designed to perform a single set of functions".[15] By describing a human being in terms of such a model, we do not exactly flatter him, but neither for that matter do we become any the more objective or scientific.

Because a machine is usually at its best when it deals with some single function in isolation, for example, the principle of the wing abstracted from the bird and given its concrete translation in the aeroplane, the most ineffective kind of machine is that which tries to give a realistic imitation of man or beast. "Technics remembers Vaucanson for his loom, rather than his life-like mechanical duck which not merely ate food but went through the processes of digestion and excretion".[16] Perhaps this is one of the reasons why the machine to enjoy strawberries and cream, or to appear friendly seems so stupid.

Turing assumes here that the fundamental characteristic of intellectual activity is to give a yes or no answer. In other words, he identifies logic with thinking and implies that intelligence and the capacity for emitting logical noises are identical. Human thought is stripped of its emotive, volitional and pragmatic characteristics, which is pretty roughly what we mean by a mechanism. Even Kenneth Craik who was very fond of drawing analogies between computing machines and human minds, appreciated that it was "illegitimate to separate thought completely from feeling",[17] Feeling and a capacity for aesthetic appreciation may play an important role not only in ordinary thinking, but also in mathematics.[18]

There are certainly methods of performing computations other than that employed by human beings: there are calculating machines, slide-rules, abacuses, etc. But it is important to distinguish between the end-result and the method by which it is arrived at. To take a simple example, the way in which a logical machine operates: the premises are fed in, all its possible combinations are developed (in the form of Boolean expansions), inconsistent alternatives are eliminated and the answer is flashed out. The whole process is simply one of classification and sorting. Insight does not come in at all.

Thinking is usually defined epistemically (or psychologically). It now seems to refer to the logical manipulation of strings of symbols, so that any set of operations resulting from the functioning of a physical bank of thermionic values [or silicon chips], which gives the correct answer, is identified with thinking. It does not, however, follow that because the end-results are identical the intervening processes are too.

One has to be careful not to identify logic with psychology. Apart from the testing of its correctness or incorrectness, thinking is no way the concern of logic. From the point of view of logic there is little to choose (except in precision) between the performance of a man and a machine. A human being might well be looked at as a machine for producing logical conclusions. The propositions are fed in through the eyes or ears (input) and after a short interval the answer emerges in words, either as speech through larynx and mouth movements, or as writing through hand movements (output). From a behavioural point of view, there is little or nothing to choose between his reaction and that of a machine. Epistemically, the picture is quite different, the process of thinking about these propositions, making inferences from them and knowing their truth or falsity, is something entirely different from the logical formulae contemplated.[19]

One might say that a logical calculus is in a sense the very antithesis of thinking, since it is a mechanical routine substituted for our intuitive and often vague and imprecise thought processes. "As a material machine is an instrument for economizing the exertion of force so a symbolic calculus is an instrument for economizing the exertion of intelligence".[20] By applying the rules of the system, we are enabled to make very long and complicated chains of deductions with a minimum of thought and effort. We translate our thoughts into the basic signs of the calculus, perform operations upon them and retranslate back again. We have produced what in effect is a simple logical machine.

But if we grant that logical machines are complex pieces of symbolism, a development of the visual aids to thinking which we have known for centuries, in order that the signs may acquire a significance they need to be given a significant logical or mathematical interpretation. As Whitehead tells us, though we can study the art of the practical manipulation of these signs without needing to assign any meaning to them, abstract calculi only posssess a serious scientific value when they can be given an important interpretation.[21]

Neglect of the pragmatic or instrumental aspects of such machines leads to the tendency to attribute to them a capacity for thinking which they only have by proxy. The transformation of formulae according to a fixed set of logical rules is not, however, a sufficient criterion of thinking. Unless the resultant formulae or strings of symbols are translated in terms of their referents, the transformation remains a meaningless array of marks. The whole argument then reduces to a tautology; such machines cannot be said to be thinking unless there is an intelligence to programme the machine and interpret the end-result, which is a form of thinking anyway.[22]

These mechanical calculi then not only need a power supply, but also an intelligence to operate them, a staff of technicians to translate mathematical problems into a form which the machines can handle. To quote Hartree, "all the thinking has to be done beforehand, by the designer and the operator who provides the operating instructions for a particular problem, all the machine can do is to follow these instructions exactly".[23]

[Even today when most people have their own PC one still has to be what is called computer-literate. Although software such as Windows is there to help us, one cannot prevent it crashing, and this will involve effort on our part to put it right. Even in the case of computer-operated machines programmed to perform specific tasks as in a car assembly plant, breakdowns may occur, and a human trouble-shooter will have to be called in to sort out the problem.]

IV. Mental processes and brain processes

A critic might recommend us to adopt a more sympathetic approach. Science as opposed to popular linguistic usage may, in a hundred years' time, verify Turing's hypothesis that thinking may be defined operationally. Popular usage, particularly in science has, as often as not, been shown to be wrong.

The behaviourist may, of course, use language in the way he pleases, as long as he restricts himself to talking about brain states or patterns of behaviour. But the difficulty arises when he attempts to show, as he is bound to do, the relevance of the behaviouristic way of talking to the way we normally talk about our sensations, feelings, volitions, etc. How such phrases as "I see red", "I feel happy", etc., are ordinarily used then becomes highly relevant, the behaviourist has to define them in terms of his own peculiar way of talking.

A.J. Ayer[24] and Mrs. M. Kneale[25] have pointed out that sentences such as "this is red", or "I am thinking", are not equivalent to the physical propositions in terms of which the behaviourist wishes to translate them. There is no logical contradiction involved, for example, in asserting the first person proposition "I feel a pain in my upper left molar", and denying the correlated third person physical proposition, "Mr. X has a carious upper left molar". The dentist after examining the tooth may deny categorically that there is anything wrong with it. One might try to save the physical explanation by saying that this is really a case of referred pain, but in that case the spatial correlation breaks down.

The definition of psychological phenomena in terms of behavioural patterns is, as C.I. Lewis[26] tells us, comparable to the physicist's assertion that a specific pitch is a particular frequency of harmonic motion. The correlation of the two, however, could never have been established if (a) "Middle C" did not first mean something identifiable without reference to vibration, and if (b) "a vibration of 256 per second", did not first mean something identifiable without reference to sound. As a result of repeated observation and experiment we come to have a high degree of inductive assurance as to their correlation, and thereafter identify pitch by the physical vibration.

But when we attempt to correlate "seeing green", "feeling pain or anger", with specific brain states or patterns of behaviour the situation is somewhat different, since the correlation is less well-established and in many points of detail quite undetermined. We cannot state with precise accuracy what specific kinds of brain states or behaviour can be correlated with "seeing green" or "feeling pain or anger". In any case the psychological phenomena themselves are not always precisely measurable.

Behaviourists sometime speak as if all they meant by "seeing green" or "suffering pain", were merely certain patterns of linguistic or bodily behaviour. It becomes clear from the above analysis that such statements are locutions for complicated inductive procedures, as they assume (a) we already know psychologically what "seeing green" or "feeling pain or anger" is like; (b) that we have been able to identify certain fluctuating behavioural patterns or brain states, which do not necesarily have the same constancy as their physical counterparts (as when we are able to correlate musical sounds with physical vibrations), and (c) that we have a large measure of inductive assurance as to their correlation, which once again is far from being the case. Behaviouristic operational definitions need then to be taken as tentative hypotheses.

[In order to update this we need to substitute cognitive scientists for behaviourists. Although the former move on a much more theoretical plane, they seem to assume that a physiological fact, such as a brain process, is as relatively simple and straightforward as our perception of the colour "red". Brain processes, however, only have a significance when there is an intelligence (i.e., a human mind) to interpret them. When the brain physiologist reports on the electrical activities of a subject's brain as given through his instrument readings, he is reporting on his own conscious interpretation of them. He then attributes these readings to the subject, to whom he may yet deny consciousness.

Knowledge about brain processes then involves complicated observational and experimental procedures, in the statement of which mathematics and logic are involved. An interesting feature of the mind-brain-dependency hypothesis is that it would make physics as well as mathematics and logic depend on brain physiology. But since the laws of brain physiology ultimately depend for their significance as well as coherence on mathematical and logical laws, we find ourselves involved in a circular argument. Whether this is a virtuous or vicious one depends on your point of view.]

The basic assumption underlying the digital computing machine analogy is that thinking operates in the form of an atomic system. It accepts the Wittgenstein (of the *Tractatus*) view of the world as a structure of atomic facts, each fact being independent of the other, and that our logic of atomic propositions reflects it. Such a view fits in naturally with Kenneth Craik's assertion that the brain is a calculating machine which is a symbolic model of the external world.[27] The progenitor is, of course, the early Wittgenstein;

"we make to ourselves pictures of fact", "The picture is a model of reality", "It is like a scale applied to reality".[28] It is but a step from the slide-rule to the calculating machine. Indeed one might say that modern digital computers are electrified pieces of Wittgensteinian Tractarian logic, i.e., of truth-tables.

The above theory of logical atomism has as its corollary that the brain is also an atomic system and functions in a purely additive fashion, that all our thought and behaviour is but the summation of the individual behaviour of its 10^9 brain cells. Lashley[29], Goldstein and Golla believe that the brain works rather as an organic system. As Golla puts it, "In fact even on the neuro-physiological level we have to regard the nervous system as an organic whole and not as an integration of reflex arcs each with an unalterable function".[30]

Defenders of the digital computing machine analogy seem implicitly to assume that the whole of intelligence and thought can be built up summatively from the warp and woof of atomic propositions. There is a good deal of psychological evidence that we think and perceive in terms of "gestalten" (or wholes), which are not merely the algebraic sums of the elements into which they may be analysed. The fact that there are equally good psychological and physiological theories to which the digital computing machine model does not apply is in this paper, at least, not dealt with by Turing.

It is not altogether too fanciful that the machine analogy together with emphasis on overt behaviour and abnegation of private experience may, when the doctrines of "Cybernetics" finally percolate down to the lower grades of the Civil Service, lead us to be regarded, more than ever before, as if we were mechanical objects. It is not such a far cry from Aristotle's view that slaves were just human tools, to some future benevolent dictatorship of the Orwell 1984 type, where men may be seen as little else but inefficient digital computers, and Big Brother as the Master Programmer.

Afterword, April 2000

What I did not deal with in my reply, nor for that matter did Turing, was the socio-historical context in which our thoughts, feelings, desires, beliefs, etc., occur, and which give them their meaning. Human beings, (which includes their minds and bodies) do not exist *en vacuo*, but are influenced by the society and its institutions which they are, to use a Heideggerian phrase, "thrown into."

Analytical philosophers too have generally failed to take the social context into account in their discussions of consciousness and mind. Steven Rose, a socially minded neuro-physiologist, has criticised their position in the following terms. "But the argument that consciousness is not merely the obverse of unconsciousness, that it is not some static brain/mind process but rather a socially, historically, developmentally engendered statement about the relationship between an individual and the surrounding world, cuts little ice – perhaps because the discussion is still being conducted primarily between philosophers trained in the classical mode and neuroscientists, and this excludes the social and historical domains".[31] Rose here echoes the young Marx's claim that we are essentially species (i.e., social) beings. Although society is not the only factor influencing the growth of consciousness, it is still an important one.

Turing also appears to be uninterested in the social context in his discussions of intelligent activity. When he asks "Can we in fifty years' time construct a machine which will play such an imitation game?", he fails to note that the very notion of a game is a social phenomenon. The rules of a game are conventions socially agreed upon by us, as in chess, football, cricket, and even in simple children's games. One can give these rules a mathematical formulation (as in the Theory of Games) but this would not make them any the less social. In postulating such a game-playing machine, Turing implicitly assumes a social context, which gives significance to the game and its rules. He further assumes that this context will be roughly the same in the year 2000 as it was in 1950. Otherwise, the two players would fail to understand each other and the game would lose its point.

Further, the social and historical factors on which, if we are to believe Kuhn, scientific progress depends, would seem to be unspecifiable in practice if not in principle. If we cannot as yet give a complete description of these factors, it is difficult to see how one could produce a programme for them to feed into a machine. Some sociobiologists would like to reduce social behaviour to a fact about our genes being programmed in a specific way. But even ants, their favoured field of study, show a marked capacity to adapt their innate behaviour to new environmental conditions. And as far as I know no gene for altruism has yet been located in the ant's chromosomes, nor, despite the mapping of the human genome, one for selfishness in man. The concept of the "selfish gene" seems little else than a form of anthropomorphism, based on a Samuel Smiles self-help philosophy and a Milton Friedman economics.[32]

Manchester Metropolitan University

References

1. A.M. Turing, "Computing Machines and Intelligence", *Mind* N.S. 59 October 1950. Edinburgh, Thomas Nelson and Son, pp.333-460.
2. W. Mays, "Can Machines Think?", *Philosophy*, April 1952, London Macmillan, pp.148-162.
3. G. Ryle, *The Concept of Mind*, London, Hutchinson 1949.
4. John R. Searle, "Consciousness, Explanatory Inversion and Cognitive Science", in *Philosophy of Psychology: Debates on Psychological Explanation*, eds. Cynthia and Graham Macdonald, Vol I. Oxford, Blackwell 1995, p.332.
5. Turing, "Computing Machines and Intelligence", p.454.

6. On unspecifiability see, Michael Polanyi, *Personal Knowledge*, London Routledge and Kegan Paul 1958, pp. 257-259
   In the Manchester seminar on "Mind and the Computing Machine" of 1949, in which both Polanyi and Turing took part, Turing answered Polanyi by saying that the mind is unspecifiable, because it has not been specified. It is a fact, he went on, that it would be impossible to find the programme inserted in quite a simple machine, and we are in the same position with regard to the brain, so that the conclusion that the mind is unspecified does not follow. But Turing can only come to this conclusion, assuming that his analogy is correct, by identifying the mind with the brain. Polanyi is not saying that the elements of the brain are unspecifiable but only those of the mind.
7. Michael Polanyi, J.R. Lucas and Roger Penrose have used Gödel's theorem to show that minds are different from machines. See my discussion of their position in "Mind and the Machine", *Journal of the British Society for Phenomenology*, January 1996, and in "Turing and Polanyi on Minds and Machines", *Appraisal*, Vol.3, No. 1, March 2000, pp. 7-8.
8. *New English Dictionary*. Oxford, Vol.VI. Part II, M-N p.7.
9. G. Jefferson. "The Mind of Mechanical Man" The Lister Oration, *The British Medical Journal*, 1949, p.1110. Jefferson was a well known Manchester neuro-surgeon who was interested in the mind-body problem.
10. Stephen Hawkins in a speech given at a private White House dinner, quoted by Joseph Margolis in "... But what about the year 3000?", *Weekend FT*, London, Dec. 4-5 1999, p.10.
11. D.R. Hartree, *Calculating Instruments and Machines*, Cambridge, Cambridge University Press 1950, p.54.
12. L.Mumford, *Technics and Civilization*, London, Routledge 1934, p.424.
13. Bertrand Russell, *A History of Western Philosophy*, London, Allen & Unwin 1946, p.856.
    Russell, with Dewey's pragmatic conception of collective power in mind, went on to say, "the concept of "truth" as something dependent on facts largely outside human control has been one of the ways philosophy has inculcated the necessary element of humility. When this check upon pride is removed, a further step is taken towards a certain kind of madness – the intoxication of power.... I am persuaded that this intoxication is the greatest danger of our time."(p.856)
14. G. Ryle, *The Concept of Mind*, p.82.
15. L. Mumford, *Technics and Civilization*, p.11.
16. *Ibid.*, p.32
17. K. Craik, *The Nature of Explanation*, Cambridge, Cambridge University Press 1943 p.96.
    Craik, a Cambridge psychologist, believed the function of scientific explanation was to make symbolic models of the world in terms of which we can make predictions about it.
18. cf. J. Hademard, *The Psychology of Invention in the Mathematical Field*, Princeton, Princeton University Press, 1945.
19. Cf. W.E. Johnson, *Logic*, Part II, Cambridge, Cambridge University Press 1922, Chapter I.
    Johnson introduced the distinction between the epistemic conditions of inference, and the constitutive ones. The former referring to the actual epistemological process of inference, the latter to the formulae contemplated. This is similar to Husserl's distinction between noesis and noema, where noesis is the intentional act and the noema the object intended.
    Peirce's views on logical machines are also of interest here. See *Peirce's Collected Papers*, Vol. II, Elements of Logic, eds. Hartshorne and Weiss, Cambridge, Harvard University Press, 1932. See 2.56,2.59.
20. W.E. Johnson, "The Logical Calculus I" *Mind* N.S. Vol.I, Edinburgh, Thomas Nelson and Son, 1993, p.3.
21. Cf. A.N. Whitehead, *Universal Algebra*, Cambridge, Cambridge UniversityPress 1898, pp.3-5 .
22. Cf., W.E. Johnson, *Logic*. Part II, Chap II, section 2, and "The Logical Calculus I", pp.3-6.
23. D.R. Hartree, *Calculating Instruments and Machines*, p.70.

24. A.J. Ayer, *The Foundations of Empirical Knowledge*, London, Macmillan 1940,p.151.
25. M. Kneale, "What is the Mind-Body Problem?", London, *Proceedings of the Aristotelian Society*, 1949-50, pp.105-22.
26. C.I. Lewis " Some Logical Considerations Concerning the Mental", in H. Feigl and W. Sellars. *Readings in Philosophical Analysis*, New York, Appleton, Century, Crofts 1949, pp.390-391.
27. Cf., K. Craik. *The Nature of Explanation*, p.57.
28. L. Wittgenstein, *Tractatus, Logico-Philosophicus*, London, Kegan Paul, Trench, Trubner and Co, 1947. See 2.1, 2.12, 2.1512.
29. For Lashley's criticism of the neo-pavlovian association theory of learning, see K.S. Lashley and M. Wade "The Pavlovian Theory of Generalisation" in *Psychological Review*, Vol. 53 1946, pp.72-86, pp.72-86.
30. F.L. Golla, *Journal of Mental Science*, 1938, 84.9, quoted from E.L. Hutton "The Relationship of Mind and Matter to Personality", in *Neuropsychiatry*, London, Lewis, 1950.
31. Steven Rose "How can meat be sentient?". *The Times Higher Educational Supplement*, May 5 2000, p.25.
32. This paper was read at the annual conference of the British Society for Phenomenology held at St. Edmund Hall, Oxford, 7-9 April 2000. My thanks are due to Tony O'Connor and Bill Hamrick who persuaded me to publish it.

Appendix

Feedback mechanisms and Mind

The limitations of the digital computer model of the mind have been recognised by Turing and others, when they endeavour to supplement it by other automatic mechanisms. Among these are negative feedback (or self-regulatory) devices which they claim can simulate human learning and purpose. Examples are to be found in gun-sighting mechanisms, automatic pilots, electric irons and kettles, and in the humble water cistern.[1]

I need therefore to make a few remarks about such mechanisms which are error-correcting. In this respect they resemble inductive rather than deductive mechanisms. One of the assumptions of this model is that trial and error behaviour is a fundamental feature of human and animal learning as well as purposive behaviour. Turing in the Manchester Seminar in answer to the question, whether one can give purpose to a machine, put it as follows "this sort of thing can be done by trial and error and the purpose is the use of previous combinations plus trial and error".[2] The equation is then simple; stored memory plus trial and error equals purpose. On this view purposive learning is a process which can be imitated mechanically (or electronically). Rat and rabbit look a likes, and even not so dumb waiters have been constructed on this model, and to make them more realistic have an inbuilt computer which can be programmed in various ways.

But we need something more than a feedback device to explain adaptive behaviour, especially at its higher levels in human beings. Human purposive learning is not just a step-by-step adjustment of previously acquired behaviour, but involves insight and the use of short-cuts. This may be seen in the solution of even simple problems. What is left out in such mechanical goal-seeking activities (no doubt as irrelevant to an objective analysis) is not only the power to anticipate the results on a conceptual level (which might be programmed), but also to evaluate our actions with reference to ideal aims and values. Because of this our responses are not predictable with the same accuracy as the responses of a simple goal-seeking mechanism. No doubt a philosophical materialist would dismiss aims and values as mere illusion, but without such notions a good deal of moral and social behaviour would become unintelligible.

Proponents of the mechanistic approach have tried to deal with the unpredictability objection as follows. They assert that what we take to be volitional activity or free will in a human being, can be simulated by an artefact containing a randomising element, and that behaviourally there is no reason why the two should be distinguishable. But the parallel between "randomness" in an artefact and free-will in a human being does not take us very far. Two series of events may exhibit randomness, but in one case it may be due to design, as when we deliberately shoot at a target so as to produce a random series of hits. In the other, it may be the outcome of a complex interplay of physical causes, as in the series of digits produced by a randomising device such as that used in the National Lottery (which has itself been designed by a human agent). As to the relation of randomness to human behaviour, although to an external observer a man's actions may appear random they immediately acquire a significance when one becomes aware of his motivation. He might, for example, be James Bond trying to put the K.G.B. off his track.

References

1. For a more detailed analysis see, W. Mays, "Probability Models and Thought and Learning Processes", *Synthese* Vol XV, No. 2, June 1963. Dordrecht, Reidel, pp.204-221.
2. From notes of a seminar on "Mind and the Computing Machine", held in the Manchester University Philosophy Department, 27 Ocober 1949, which I attended. For a summary of this seminar see, W. Mays "Turing and Polanyi on Minds and Machines", *Appraisal,* Vol.3, No.1, March 2000, pp.7-8.