CAN MACHINES THINK?

W. MAYS, Ph.D.

I

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."

Apparently 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 an imitation 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 computors to play the imitation game. He therefore replaces the original question, "Can machines think?" by the question, "Are there imaginable digital computors which would do well in the imitation game?"

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. Nevertheless, 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 shapeless masses of metal cogitating in a Rodin-like manner.

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

It is therefore of some importance, as well as interest, to know how people actually use the words to-day, and not how they may use them at some future date. As we are unable to embark upon a Gallup 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.

The most relevant of the O.E.D. definitions of a machine 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 and uniformity of a machine."[1]

From the point of view of modern neo-behaviourism this account may seem a little old-fashioned. Perhaps it is as 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 to-day. 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 reverie 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,"[2] and 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 streamline 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 conation which we assign to a human being. We have expressly ruled out any internal private life. Though it may duplicate our overt or external behaviour, it cannot, we say, 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 refuse to 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 indulging in linguistic skills), whereas we would

[1] *New English Dictionary*. Oxford, Vol. VI, Part II, M–N, p. 7.

[2] G. Jefferson, *The Mind of Mechanical Man*, The Lister Oration. *The British Medical Journal* (1949), p. 1,110.

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

still be inclined to attribute some sort of mental life even to a moron.

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 conation are inextricably intermingled. One of the few places where pure thought is to be found is in Immanuel Kant's *Critique*.

If, however, we did come across an artifice which showed every sign of intelligent behaviour, and was yet in appearance very unlike a man, 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 (a man from Mars) or some sort of mechanism. Biologists are familiar with such situations.

Consider an illustrative example. To-day there is little risk of confusion between a motor-car and a motor-cycle. However, some enterprising person may yet produce a gyroscopically driven vehicle intermediate between the two. And in practice we would introduce a new class-label to cover this case.

So, in the same way, it may be necessary to introduce a new label 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 Karel Capek, we call them "robots," those devices which fall in the twilight zone between man and the normal run of machines; devices which it must be remembered are still in the realm of fiction—in the imagination of their authors.

In this connection it might be a good thing to drop the word "machine," with its emotional overtones of clanging metal, and use some such neutral word as "artifice." Machines which can perform logical and mathematical operations are very different from the steam-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 conceptions of minds and machines. At least for the unphilosophical person, 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

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

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. The number of facilities a sufficiently fertile imagination could project into these contrivances appears to have no upper limit.

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 produced by natural methods and artificial men (or robots). The same shame might be attached to be being born the natural way as we nowadays attach to illegitimacy. A different label might indeed become a social necessity.

In any case, the words "Thinking" and "Intelligence" are vague, replete with ambiguities and intrinsically subjective in character. The underlying assumption of many modern theories of intelligence is that intelligence is a simple quality and directly measurable. There are good grounds, however, for believing otherwise, that it is heterogeneous in character, and only measurable by indirect methods.

II

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 part of the quoted passage:

"Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain." (p. 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 therefore not exclusive and Jefferson need not accept either.

Further, if I understand Jefferson rightly, he is not just 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

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

mathematical theorems, because of thoughts and emotions felt: *felt* being the operative word here. Turing apparently 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 person who bears the strongest resemblance to digital computors, are the so-called lightning calculators, who are often uneducated men, and even sometimes feeble-minded.

Turing's sonnet-writing machine (p. 446), which shows up so well in the *viva-voce* examination, is not I think relevant to the discussion. It does not follow that because q, the sonnet-writing behaviour, occurs, that it is due to p, thoughts and feelings felt; it may be due to a bank of relays or a regiment of monkeys hammering away on typewriters, though their speed and output would be lavishly increased if they were armed with modern digital computors. Jefferson could still refuse to accept the imitation game as a definition of thinking, if it was not due to p.

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. It is a deterministic machine language in the grand manner of nineteenth-century Newtonian physics.

In the psychological language used by Jefferson words like consciousness, freewill, decision, etc., appear. But words like desiring and feeling ought by their very nature to be prohibited from Turing's account. And though it has been assumed by Carnap and others that the psychological language is translatable in terms of the physical, there is as we shall see little justification for this belief.

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 language, otherwise we shall find our account infected by more than a mild attack of animism.

Turing, in his account of the wonders of digital computors, 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

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

to understand and speak English. 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." (p. 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. It would be interesting to know whether the subjection of the machine to punishments and rewards consists in anything more than the switching on and off of its power supply. The use of such emotively toned words (which also seem to express value judgments) makes one think immediately of someone precariously balancing a calculating machine on his knee and chastising it.

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 which takes the form.

"I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do *X*."

It will be observed that *X* is generally an item from the psychological language, or a value judgment.

"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 make one 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 worth while making such a machine, would it not be another marvel in the thesaurus of the mechanical necromancer? Or does Turing implicity recognize that it would be meaningless to talk about a machine "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. As Hartree 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."[1]

[1] D. R. Hartree, *Calculating Instruments and Machines*, p. 54 (Cambridge 1950).

III

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

"Is it true that by modifying this computor to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, *C* can be made to play satisfactorily the part of *A* in the imitation game, the part of *B* being taken by a man?" (where *C* is, of course, the digital computor) (p. 442).

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

In principle there is perhaps no reason why an elephant should not have wings, there is no logical reason why we should not all live to be as old as Methuselah, or for survival, but there may be important physical reasons to the contrary. There may be important physical limitations, spatial, temporal, and mechanical, why it is not possible. I believe, but I may be mistaken in this, that engineers are already finding that there is an optimum limit to 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 in those journals devoted to astounding science fiction.

Mumford points out that certain machines have already reached the limit of their development, e.g. the printing press, the water turbine and even the telephone system, the only gain is one of cheapness and universality. If a very good page of print is needed it still has to be set up by hand; "there are bounds to mechanical progress within the nature of the physical world itself. It is only by ignoring these limiting conditions that a belief in the automatic and inevitable and limitless expansion of the machine can be retained."[1]

Turing tells his readers that digital computors "can in fact mimic the actions of a human computor very closely" (p. 438). His statement of fact seems to be a most dubious one, unless he is using mimicry in a very peculiar sort of way where it can have no reference to "purposive imitation." As a large part of his case

[1] L. Mumford, *Technics and Civilization*, p. 424 (Routledge 1934).

rests on his conception of "mimicry" it is necessary to give an analysis of it.

Apparently it consists of three stages.

1. Asking a human computor 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 I^s and O^s.

3. Inserting the programme into the machine.

"Mimicking" for Turing is then a complex operation and consists in a human programmer translating the human computor's 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 computor" seems to be used in a rather ambiguous way. Sometimes it refers to the machine standing on its own feet as it were, and sometimes to the loaded machine with a programme inserted in it. These are two very different things. Even if we are generous and take the latter interpretation, and even if we conveniently erase from our memories that it has been programmed by the "slave to the machine," it still could not be said to be mimicking a human computor, as mimicking means purposive imitation. If Turing wishes to restrict himself purely 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 relationship of similarity between its behaviour and the behaviour of a human computor, and not that the relationship is one of "mimicking." ["Mime" in its original use as seen, e.g. in ballet, was rather an attempt to communicate emotive states by overt symbolism.]

However, there are limitations in the art of translation. Shakespeare loses his essential quality when translated into basic English. Even natural languages owing to differences in physical structure and vocabulary are not precisely translatable into each other—"traduttore traditore." If basic English is thus limited what are we to say about Boolean algebra, a language with an alphabet of two letters, 1 and 0, and an extremely simple grammar? Despite Leibnitz's auguries, as a universal language it is hopelessly inadequate. Such a view naturally accepts the principle of extensionality that all logical expressions are translatable into terms of truth-values (i.e. 1^s and 0^s) and, surprisingly enough, implies that semantic expressions are also thus reducible.

https://doi.org/10.1017/S003181910002266X Published online by Cambridge University Press

It is merely a conceit of the logician to imagine that the main function of natural language is to transmit information rather than to communicate personal and social feeling. In poetry, literature and everyday discussion, the truth and falsity of statements quite often play a very minor and unessential rôle. One might say that verse written in abstract logical form, with signs expressing combinations of truth-values, would not be poetry, nor even dull prose, but just bad logic.

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 resembles 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. p. 456, l.2 and l.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 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.

On the face of it, Turing's account bears a strong resemblance to the "tabula rasa" theory of 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 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-mind may have a good deal more internal structure than Turing bargains for.

What, on Turing's view, corresponds to the human mind is, however, not just the machine, but the machine plus the instructions 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 then contains a selective 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 statistical concepts. The abolition 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 few natural machines in nature; "inventing machines is not copying things found in inanimate Nature."[1] If we want to find examples in nature of "self-maintaining routine observing systems," we have rather to look to living organisms. The machine is not then something superior to the men and women who design and construct it, it is, in fact, an inferior type of animal, "a sort of minor organism, designed to perform a single set of functions."[2] 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.

And because a machine is usually at its best when it deals with some single function in isolation, e.g. 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 for his life-like mechanical duck which not merely ate food but went through the routine of digestion and excretion."[3] Undoubtedly this is one of the reasons why the machine to enjoy strawberries and cream, or to be friendly, appears so stupid.

Turing assumes in his article 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, conative and pragmatic characteristics, which is pretty roughly what we mean by a mechanism. Even Craik, who was very fond of drawing analogies between calculating machines and human minds, nevertheless appreciated that it was "illegitimate to separate thought completely from feeling."[4] Feeling and a capacity for aesthetic appreciation play an important rôle not only in ordinary thinking, but also in mathematics.[5]

There are certainly methods of performing computations other than that employed by human beings; there are calculating machines, slide-rules, abacus, 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

[1] G. Ryle, *The Concept of Mind*, p. 82 (Hutchinson's 1949).
[2] L. Mumford, *Technics and Civilization*, p. 11 (Routledge 1934).
[3] L. Mumford, *Technics and Civilization*, p. 32 (Routledge 1934).
[4] K. Craik, *The Nature of Explanation*, p. 86 (Cambridge 1943).
[5] Cf. J. Hadamard, *The Psychology of Invention in the Mathematical Field* (Princeton 1945).

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.

With suitable adaptations a machine could be constructed with the facility for performing intelligence tests, e.g. Raven's matrices, at a much greater speed than little boys and girls could. One might even talk of the I.Q of the machine, yet no one would seriously think of attributing intelligence to it. What is important is not what it does, but how it does it: a point overlooked by intelligence testers in the past.

Thinking is usually defined epistemically (i.e. psychologically). It now seems to refer to the logical manipulation of strings of symbols, so that any process such as a physical collection of relays, or a bank of electronic valves 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. If a man picks the Derby winner by sticking a pin into the back page of the *Sporting Chronicle* it does not follow that he has been exercising conscious judgment, that he has been a diligent student of Oakeshott and Griffith's *Guide to the Classics*.

One has to be careful not to identify logic with psychology. Apart from the testing of its correctness or incorrectness, thinking is in 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 ears (input) and after a short pause the answer emerges from the mouth (output) in the form of words. Behaviouristically, 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, of knowing their truth and falsity, is something entirely different from the logical formulae contemplated.[1]

Indeed, 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."[2] By applying the rules of the system we

[1] Cf. W. E. Johnson, *Logic*, Part II, Chapter I (Cambridge 1922).
Pierce's views on logical machines are of interest. *C. S. Pierce's Collected Papers*, Vol. II, *Elements of Logic*, 2.56, 2.59 (Harvard 1932).

[2] W. E. Johnson, "The Logical Calculus I" *Mind*, N.S., Vol. I (1892), p. 3.

are enabled to make very complicated and long chains of deductions with a minimum of thought and effort. We translate our concepts 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 specific logical or mathematical interpretation. As Whitehead tells us, though we can study the art of practical manipulation of these signs without needing to assign any meaning to them, abstract calculi only possess a serious scientific value when they can be given an important interpretation.[1]

Neglect of the pragmatic or instrumental aspects of such machines, leads to the tendency to attribute to them a capacity for thinking which they have only 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 patterns of symbols are retranslated 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.[2]

These mechanized calculi not only need a power supply, but also an intelligence to operate them, a staff of mathematicians 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."[3]

IV

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 can 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

[1] Cf. A. N. Whitehead, *Universal Algebra*, pp. 3–5 (Cambridge 1898).

[2] Cf. W. E. Johnson, *Logic*, Part II, Chapter III, § 2. "The Logical Calculus I," *Mind*, N.S., Vol. 1 (1892), pp. 3–6.

[3] D. R. Hartree, *Calculating Instruments and Machines*, p. 70 (Cambridge 1950).

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, the relevance of the behaviourist 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," are ordinarily used then becomes highly relevant, the behaviourist has to define them in terms of his own peculiar way of talking.

Further, Ayer, Mrs. Kneale[1] and others have pointed out, that sentences like "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, e.g. in asserting "I feel a pain in my left molar," and denying the correlated 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.

The definition of psychological phenomena in terms of behavioural patterns is, as C. I. Lewis[2] points out, comparable to the physicists' 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 any degree of accuracy what specific kinds of behaviour are correlated with "seeing green" or "feeling pain or anger." Discussions of physical methods of measuring psychological phenomena usually lead to an impasse.

Behaviourists sometimes 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 pro-

[1] Cf. A. J. Ayer, *The Foundations of Empirical Knowledge*, p. 151 (Macmillan 1940).

M. Kneale: What is the Mind-Body Problem? *Proceedings of The Aristotelian Society* (1949–50), pp. 105–22.

[2] C. I. Lewis, "Some Logical Considerations Concerning the Mental," H. Feigl and W. Sellars, *Readings in Philosophical Analysis*, pp. 390–91 (Appleton–Century–Crofts 1949).

cedures, 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, which do not have the same constancy nor even objectivity as their physical counterparts, 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 very tentative hypotheses.

The relationship between sentences in the psychological (or ψ) language, and sentences in the behaviouristic (or β) language is then not one of entailment, the most we can say is that there is a probability relationship between them. The two series of sentences $\psi_1, \psi_2, \psi_3, \ldots \psi_n$. $\beta_1, \beta_2, \beta_3, \beta_4, \ldots \beta_n$ seem rather to be related together as elements are in a frequency distribution, or, to put it more precisely, there is merely a statistical dependency. It is also doubtful whether we can use the word correlation here. Correlation in its strict sense suggests that the relationship is a quantitative one, and only perhaps in simple "stimulus-sensation" situations can it be shown that this dependence is in any way quantitative. Statements in the ψ language cannot therefore be translated into statements in the β language. Otherwise the correlation between them would be 1, which is patently not the case.

The basic assumption in applying the calculating machine analogy to the mind is that thinking operates in the form of an atomic system. It accepts Wittgenstein's 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 of the world and knowledge fits in naturally with Craik's assertion that the brain is a calculating machine which is a symbolic model of the external world.[1] The progenitor is, of course, Wittgenstein; "We make to ourselves pictures of fact." "The picture is a model of reality." "It is like a scale applied to reality."[2] It is but a step from the slide-rule to the calculating machine. Indeed one might say that modern digital computors are electrified pieces of Wittgensteinian logic.

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, Goldstein and Golla believe that the brain works rather as an organic system. "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

[1] Cf. K. Craik, *The Nature of Explanation*, p. 57 (Cambridge 1943).

[2] L. Wittgenstein, *Tractatus Logico-Philosophicus*, 2.1, 2.12, 2.1512 (Kegan Paul 1947).

reflex arcs each with an unalterable function."[1] Defenders of the 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," 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 computing machine analogy does not apply, seems to have been silently passed over by Turing.

To summarize, there are doubts whether the principles of atomicity and extensionality apply in logic. In the light of the work of the Gestalt psychologists it would seem that we think in terms of gestalten. In physiology too, there is a good deal to be said in favour of the thesis that the nervous system operates as an organic whole. Against this background the analogy drawn by Turing between brains, thought processes and calculating machines seems a rather forced one, and begs more questions than it solves.

It is not altogether too fanciful to suppose that the machine analogy, with its 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 computors and God as the Master Programmer.

[1] F. L. Golla (1938) *Journal Mental Science* 84.9 quoted from E. L. Hutton "The Relationship of Mind and Matter to Personality," "Perspectives in Neuropsychiatry," p. 166 (Lewis 1950).

For Lashley's criticisms of the neo-pavlovian association theory of learning, K. S. Lashley and M. Wade. "The Pavlovian Theory of Generalization," *Psychological Review*, Vol. 53 (1946), pp. 72–86.

University of Manchester.