Can Digital Computers Think?

Alan M. Turing - 15 May 1951

OCR'd and edited version of http://www.turingarchive.org/browse.php/B/5

Digital computers have often been described as mechanical brains. Most scientists probably regard this description as a mere newspaper stunt, but some do not. One mathematician has expressed the opposite point of view to me rather forcefully in the words 'It is commonly said that these machines are not brains, but you and I know that they are'. In this talk I shall try to explain the ideas behind the various possible points of view, though not altogether impartially. I shall give most attention to the view which I hold myself, that it is not altogether unreasonable to describe digital computers as brains. A different point of view has already been put by Professor Hartree.

First we may consider the naive point of view of the man in the street. He hears amazing accounts of what these machines can do: most of them apparently involve intellectual feats of which he would be quite incapable. He can only explain it by supposing that the machine is a sort of brain, though he may prefer simply to disbelieve what he has heard.

The majority of scientists are contemptuous of this almost superstitious attitude. They know something of the principles on which the machines are constructed and of the way in which they are used. Their outlook was well summed up by Lady Lovelace over a hundred years ago, speaking of Babbage's Analytical Engine. She said, as Hartree has already quoted, 'The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform.' This very well describes the way in which digital computers are actually used at the present time, and in which they will probably mainly be used for many years to come. For any one calculation the whole procedure that the machine is to go through is planned out in advance by a mathematician. The less doubt there is about what is going to happen the better the mathematician is pleased. It is like planning a military operation. Under these circumstances it is fair to say that the machine doesn't originate anything.

There is however a third point of view, which l hold myself. I agree with Lady Lovelace's dictum as far as it goes, but l believe that its validity depends on considering how digital computers are used rather than how they could be used. In fact I believe that they could be used in such a manner that they could appropriately be described as brains. I should also say that 'If any machine can appropriately be described as a brain, then any digital computer can be so described'.

This last statement needs some explanation. It may appear rather startling, but with some reservations it appears to be an inescapable fact. It can be shown to follow from a characteristic property of digital computers, which I will call their universality. A digital computer is a universal machine in the sense that it can be made to replace any machine of a certain very wide class. It will not replace a bulldozer or a steam-engine or a telescope, but it will replace any rival design of calculating machine, that is to say any machine into which one can feed data and which will later print out results. In order to arrange for our computer to imitate a given machine it is only necessary to programme the computer to calculate

what the machine in question would do under given circumstances, and in particular what answers it would print out. The computer can then be made to print out the same answers.

If now some particular machine can be described as a brain we have only to programme our digital computer to imitate it and it will also be a brain. If it is accepted that real brains, as found in animals, and in particular in men, are a sort of machine it will follow that our digital computer suitably programmed, will behave like a brain.

This argument involves several assumptions which can quite reasonably be challenged. I have already explained that the machine to be imitated must be more like a calculator than a bulldozer. This is merely a reflection of the tact that we are speaking of mechanical analogues of brains, rather than of feet or jaws. It was also necessary that this machine should be of the sort whose behaviour is in principle predictable by calculation. We certainly do not know how any such calculation should be done, and it was even argued by Sir Arthur Eddington that on account of the Indeterminacy Principle in Quantum Mechanics no such prediction is even theoretically possible.

Another assumption was that the storage capacity of the computer used should be sufficient to carry out the prediction of the behaviour of the machine to be imitated. lt should also have sufficient speed. Our present computers probably have not got the necessary storage capacity, though they may well have the speed. This means in effect that it we wish to imitate anything so complicated as the human brain we need a very much larger machine than any of the computers at present available. We probably need something at least a hundred times as large as the Manchester Computer. Alternatively of course a machine of equal size or smaller would do if sufficient progress were made in the technique of storing information.

It should be noticed that there is no need for there to be any increase in the complexity of the computers used. If we try to imitate even more complicated machines or brains we must use larger and larger computers to do it. We do not need to use successively more complicated ones. This may appear paradoxical, but the explanation is not difficult. The limitation of a machine by a computer requires not only that we should have made the computer, but that we should have programmed it appropriately. The more complicated the machine to be imitated the more complicated must the programme be. This may perhaps be made clearer by an analogy. Suppose two men both wanted to write their autobiographies, and that one had had an eventful life, but very little had happened to the other. There would be two difficulties troubling the man with the more eventful life more seriously than the other. He would have to spend more on paper and be would have to take more trouble over thinking what to say. The supply of paper would not be likely to be a serious difficulty, unless for instance he were on a desert island, and in any case it could only be a technical or a financial problem. The other difficulty would be more fundamental and would become more serious still if he were not writing his life but a work on something he knew nothing about, let us say about family life on Mars. Our problem of programming a computer to behave like a brain is something like trying to write this treatise on a desert island. We cannot get the storage capacity we need in other words we cannot get enough paper to write the treatise on, and in any case we don't know what we would write down if we had it. This is a poor state of affairs, but, to continue the analogy, it is something to know how to write, and to appreciate the fact that most knowledge can be embodied in books.

In view of this it seems that the wisest ground on which to criticise the description of digital computers as 'mechanical brains' or 'electronic brains' is that, although they might be programmed to behave like brains, we do not at present know how this should be done. With this outlook I am in full agreement. It

leaves open the question as to whether we will or will not eventually succeed in finding such a programme. I, personally, am inclined to believe that such a programme will be found. I think it is probable for instance that at the end of the century it will be possible to programme a machine to answer questions in such a way that it will be extremely difficult to guess whether the answers are being given by a man or by the machine. I am imagining something like a viva-voce examination, but with the questions and answers all typewritten in order that we need not consider such irrelevant matters as the faithfulness with which the human voice can be imitated. This only represents my opinion; there is plenty of room for others.

There are still some difficulties. To behave like a brain seems to involve free will, but the behaviour of a digital computer, when it has been programmed is completely determined. These two facts must somehow be reconciled, but to do so seems to involve us in an age-old controversy that of 'free will and determinism'. There are two ways out. It may be that the feeling of free will which we all have is an illusion. Or it may be that we really have got free will, but that there is no way of telling from out behaviour that this is so. In the latter case, however well a machine imitates a man's behaviour it is to be regarded as a mere sham. I do not know how we can ever decide between these alternatives but whichever is the correct one it is certain that a machine which is to imitate a brain must appear to behave as if it had free will, and it may well be asked how this is to be achieved. One possibility is to make its behaviour depend on something like a roulette wheel or a supply of radium. The behaviour of these may perhaps be predictable, but if so we do not know how to do the prediction.

It is, however, not really even necessary to do this. It is not difficult to design machines whose behaviour appears quite random to anyone who does not know the details of their construction. Naturally enough the inclusion of this random element, whichever technique is used, does not solve our main problem, how to programme a machine to imitate a brain, or as we might say more briefly, if less accurately, to think. But it gives us some indication of what the process will be like. We must not always expect to know what the computer is going to do. We should be pleased when the machine surprises us, in rather the same way as one is pleased when a pupil does something which he had not been explicitly taught to do.

Let us now reconsider Lady Lovelace's dictum. 'The machine can do whatever <u>we know how to order it</u> to perform'. The sense of the rest of the passage is such that one is tempted to say that the machine can <u>only</u> do what we know how to order it to perform. But I think this would not be true. Certainly the machine can only do what we <u>do</u> order it to perform, anything else would be a mechanical fault. But there is no need to suppose that, when we give it its orders we know what we are doing, what the consequences of these orders are going to be. One does not need to be able to understand how these orders lead to the machine's subsequent behaviour, any more than one needs to understand the mechanism of germination when one puts a seed in the ground. The plant comes up whether one understands or not. If we give the machine a program which results in its doing something interesting which we had not anticipated I should be inclined to say that the machine <u>had</u> originated something, rather than to claim that its behaviour was implicit in the programme, and therefore that the originality lies entirely with us.

I will not attempt to say much about how this process of 'programming a machine to think' is to be done. The fact is that we know very little about it, and very little research has yet been done. There are plentiful ideas, but we do not yet know which of them are of importance. As in the detective stories, at the beginning of the investigation any trifle may be of importance to the investigator. When the problem has been solved, only the essential facts need to be told to the jury. But at present we have

nothing worth putting before a jury. I will only say this, that I believe the process should bear a close relation to that of teaching.

I have tried to explain what are the main rational arguments for and against the theory that machines could be made to think, but something should also be said about the irrational arguments. Many people are extremely opposed to the idea of a machine that thinks, but I do not believe that it is for any of the reasons that I have given, or any other rational reason, but simply because they do not like the idea. One can see many features which make it unpleasant. If a machine can think, it might think more intelligently than we do, and then where should we be? Even if we could keep the machines in a subservient position, for instance by turning off the power at strategic moments, we should, as a species, feel greatly humbled. A similar danger and humiliation threatens us from the possibility that we might be superseded by the pig or the rat. This is a theoretical possibility which is hardly controversial, but we have lived with pigs and rats for so long without their intelligence much increasing, that we no longer trouble ourselves about this possibility. We feel that if it is to happen at all it will not be for several million years to come. But this new danger is much closer. If it comes at all it will almost certainly be within the next millenium. It is remote but not astronomically remote, and is certainly something which can give us anxiety.

It is customary, in a talk or article on this subject, to offer a grain of comfort, in the form of a statement that some particularly human characteristic could never be imitated by a machine. It might for instance be said that no machine could write good English, or that it could not be influenced by sex appeal or smoke a pipe. I cannot offer any such comfort, for I believe that no such bounds can be set. But I certainly hope and believe that no great efforts will be put into making machines with the most distinctively human, but non-intellectual characteristics such as the shape of the human body. It appears to me to be quite futile to make such attempts and their results would have something like the unpleasant quality of artificial flowers. Attempts to produce a thinking machine seem to me to be in a different category. The whole thinking process is still rather mysterious to me, but I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves.