Alan Turing and the Turing Test

by Andrew Hodges

This article was requested for the forthcoming volume, *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer,* ed. Robert Epstein. I submitted this text in April 2002. It was slightly revised in accordance with the editors' suggestions in October 2002.

1. INTRODUCTION

The short and extraordinary life of the British mathematician Alan Turing involves the foundations of mathematics, the origin of the computer, the secret cryptological war against Nazi Germany, and much more. For the modern public, one of the most frequent mentions of his name comes in yet another context: the philosophy of Mind and Artificial Intelligence. Specifically, he is immortalised in the 'Turing test' for intelligence, which Turing himself called 'the imitation game'.

The famous Test appeared in Turing's paper *Computing Machinery and Intelligence,* published in October 1950 in the philosophical journal *Mind* (Turing 1950). Turing was then employed at Manchester University, where the world's first stored-program computer had been working since June 1948. Turing had not been appointed to produce philosophical papers; his primary function was to create and manage the first software for the computer. However he had also continued his research in mathematics, and had been drawn into discussion with the scientific philosopher Michael Polanyi, who encouraged him to publish his views on what Turing called 'intelligent machinery'. The point of his paper, of which the imitation game was only a part, was to argue that intelligence of a human level could be evinced by a suitably programmed computer. The imitation game lent definiteness to the idea of being as intelligent as a human being.

Turing's 1950 paper did not arise in isolation, and the purpose of this biographical sketch is to set Turing's test in the context of his life and work. The 1950 paper was an important summary of his views for the general philosophical public, but he had been developing those views for many years. It would also be a mistake to think of Turing as a mathematician making a detached comment on the potential of computers. He was very fully engaged in the development of modern computer science both in theory and in practice.

2. THE TURING MACHINE

Indeed the 1950 paper had itself an important autobiographical element, although Turing did not emphasise its personal aspect. Much of the early part of the paper involved an exposition of the concept of *computability.* The definition of computability was Turing's own great achievement of the pre-war period, when he was a young Fellow of King's College, Cambridge University. In 1936, when he was only 23, he announced the definition of what immediately became known as the *Turing machine,* thus giving a precise and convincing mathematical definition of an 'effective method.' In Turing's paper *On computable numbers, with an application to the Entscheidungsproblem* (Turing 1936-7), he gave a discussion of his definition, arguing successfully that it encompassed the most general possible process for computing a number.

By so doing, he satisfactorily answered Hilbert's decision problem for the provability of mathematical theorems. The paper did more: it also defined the concept of *universal* Turing machine and hence the principle of the modern stored-program computer. The universal machine was also an important element in the theory Turing needed to explain in his 1950 paper: the point is that all Turing machines can be thought of as programs for a single universal machine. Most striking, perhaps, is that the Turing machine, formulated fourteen years before the 'Turing Test', was also based on a principle of imitation. The 'machine' was modelled by considering what a human being could do when following a definite method. According to Turing's argument it encompassed everything that could be done by a human calculator with finite equipment, but allowed an unlimited supply of paper to write on (formalised into a paper tape marked off in squares) and unlimited time.

The basis in human calculation is emphasised in Turing's arguments. The 'squares' of the Turing machine 'tape,' for instance, originated in Turing's explanation as the squares of a child's exercise book. The atomic operations of scanning, marking, erasing, moving to left and right were likewise related to human actions. Above all the finite number of 'configurations' of a Turing machine were related to the finite number of states of mind, or finite memory, of a human calculator. This very bold appeal to modelling 'states of mind' by states of a machine seems already to anticipate the thesis of machine intelligence in 1950. Should we then say that Turing in 1950 was only re-stating the implications of what he had claimed in 1936?

A simple reading of the story would support this view. It might be argued that from 1936 onwards Turing steadfastly sought and found ways to implement his theory in practice. In 1937 Turing began a practical interest in electromagnetic relays to implement logical operations, an interest very different from anything expected of a Cambridge pure mathematician (Hodges 1983, p. 139). After 1939 this interest turned into one of immense practical importance. Turing's ingenious logic was translated into working electromagnetic machinery at Bletchley Park, Buckinghamshire, the centre of British code-breaking operations. His algorithm for breaking the Enigma-enciphered German messages, as embodied in the British 'Bombe', lay at the centre of its success. He personally headed the work of deciphering German naval messages, and led the development of many methods and machines of astonishing efficiency. He was introduced to American work at the highest level, and to the most advanced electronic technology. In this way he learnt that electronic storage and electronic circuits could make an effective and fast practical version of the 'paper tape' and configurations needed for a universal machine. He learnt electronics for himself, again a highly unconventional step. Turing emerged from the war in 1945 full of enthusiasm for engineering a practical version of the universal machine — in modern parlance a stored-program computer. By a remarkable sequence of events Turing was taken on by the National Physical Laboratory, London, with exactly this commission. His plan was submitted in March 1946 (Turing 1946). As well as pioneering key ideas in computer hardware and software design, it mentioned the idea of the machine showing 'intelligence,' with chess-playing as a paradigm. This germ then rapidly developed into the programme set out in the 1950 paper.

3. INTELLIGENCE AND INTUITION

Although basically correct, a subtle adjustment to this basic story is required, and it is one that casts light on the structure and content of the 1950 paper. A reading of that paper will show that Turing was highly aware of the natural objection that machines cannot do those things which are by nature non-mechanical: those human actions that require initiative, imagination, judgment, cause surprise, are liable to errors, and so forth. Much of his argument is directed to the claim that machines would, in fact, be capable of all these apparently non-mechanical things.

But there is no reflection of this claim in his 1936 work. In (Turing 1936-7) the 'states of mind' that Turing considered were only those employed when the mind is occupied on performing a definite method or process. There was no reference to imagination or initiative in the 'states of mind' of *On computable numbers.*

So we can ask: at what point in his biography did Turing adopt the idea that computable operations would encompass everything normally called 'thinking' rather than only 'definite methods'? Or equivalently, when did he first consider that operations which are in fact the workings of predictable Turing machines, could nevertheless appear to the human observer as having the characteristics of genuine intelligence and creativity?

Turing wrote very little about his own intellectual development, and his writings do not give a direct answer to this question. However there are two important stages in his work, not mentioned in the above account, which when considered in their context suggest a plausible answer: namely at some point after 1938 and probably in about 1941.

During the two years Turing spent at Princeton, from 1936 to 1938, he was investigating the logic of the *uncomputable.* Turing's exposition (Turing 1939) described the 'formulae, seen intuitively to be correct, but which the Gödel theorem shows are unprovable in the original system.' In this pre-war period, Turing apparently left open the possibility that the mind had an 'intuitive' power of performing uncomputable steps beyond the scope of the Turing machine (a thesis, incidentally, that was always held by Gödel himself.)

But in the period around 1941 when the immediate crisis of the Enigma problem was resolved, Turing began to discuss with colleagues at Bletchley Park the possibility of machines playing chess (Hodges 1983, p. 213). Chess-playing was in fact a civilian analogue of what they were doing in their secret work, in which mechanical methods of great sophistication were outdoing some aspects of human intuition. It seems, taking the view expressed in (Hodges 1997, 2002), that Turing probably decided in about 1941 that the scope of computable operations was in fact sufficient to account for those mental operations apparently 'non-mechanical' by the standards of ordinary language, and even the apparently uncomputable operations of truth-recognition.

There was possibly another wartime influence on his development: Turing' general exposure to modern ideas such as the neural physiology of the brain and the behaviourist model of the mind. McCulloch and Pitts (1943) related their logical model of neurons to Turing's computability; Turing returned the compliment by referring to this work. Turing was developing the picture of the brain as a *finite discrete state machine.* In a sense this was only a small step from the 'finitely many states of mind' of 1936. But it went further because Turing's postwar idea was that *all* mental function of the brain could be accommodated in this model, and not just those of a mind following a definite rule. As we shall see, Turing framed an argument to explain how this mechanical picture of the brain could be reconciled with the counter-arguments from Gödel's theorem.

Very possibly it was this new conviction that made him so enthusiastic in the closing stages of the war about the prospect of building an electronic version of the universal machine. He was not highly motivated by building a computer to work out programmed mathematical operations. His interest was more in the nature of the mind. Informally, when speaking of his computer plans in 1945, he called them plans for 'building a brain.'

4. INTELLIGENT MACHINERY

With this in mind, we can examine in more detail his very first written mention of 'intelligent' machinery in (Turing 1946). One should first note how remarkable it was, that Turing should put a speculative claim about intelligence in a purely technical, practical report. However this was entirely typical of his *modus operandi*. One should next appreciate that Turing always relished the paradox, even apparent contradiction in terms, involved in speaking of 'intelligent' machinery. First he explained how the computer could be programmed to calculate chess moves. He continued, underlining the paradox:

> This... raises the question 'Can a machine play chess?' It could fairly easily be made to play a rather bad game. It would be bad because chess requires intelligence. We stated... that the machine should be treated as entirely without intelligence. There are indications however that it is possible to make the machine display intelligence at the risk of its making occasional serious mistakes. By following up this aspect the machine could probably be made to play very good chess.

This mysterious reference to 'mistakes', which could have made no sense to anyone reading this report, is explained in a talk of February 1947 (Turing 1947). Here the idea of 'mistake-making' appeared in the context of the objection to the prospect of machine intelligence posed by Gödel's theorem. This objection (which appears as 'The Mathematical Objection' in the 1950 paper) is that no Turing machine (i.e. computer program) can do as well as a human being. The human being can see the truth of mathematical assertions which cannot be proved by following rules based on formal axioms.

Turing's post-war argument (the point of view he probably arrived at in about 1941) is, however, that human beings do not reliably see the truth of such statements. Mathematicians, their brains being discrete state machines, can only employ an algorithm. Gödel's theorem tells us that no algorithm can coincide in every case with truth-seeing, and so the algorithm is bound sometimes to fail. But if it is accepted that the mathematician is not infallible, and will sometimes fail, it follows that machines — also implementing algorithms, and therefore also making mistakes — may do equally well. To illustrate the theme of doing equally well, Turing appealed to the concept of 'fair play for machines.' This concept was essentially the idea of the imitation game. The 1950 scenario merely added dramatic detail. Thus, the imitation game had its origins in the wartime debate in Turing's own mind about how to reconcile Gödel's theorem and the apparently non-mechanical actions of human minds with the discrete state machine model of the brain.

After 1947 Turing continued to a wider and more constructive discussion of how machines might perform apparently non-mechanical tasks: how completely unintelligent micro-operations might add up to intelligent processes. It was presented in an internal report 'Intelligent Machinery' for the National Physical Laboratory (Turing 1948). This was not published (until 1968), but was in many ways the basis of his better-known and less technical 1950 exposition. One interesting feature of this 1948 report is the evidence it gives of a wartime inspiration for his new ideas. Turing referred to images of the writer Dorothy Sayers, to illustrate the commonly accepted meaning of 'mechanical' behaviour. The book he quoted was one he was reading at Bletchley Park in 1941. Turing also tellingly described 1940 as the date after which machines were no longer restricted to 'extremely straightforward, possibly even to repetitious, jobs.' He must have had his own Enigma-breaking Bombe, and other highly sophisticated codebreaking operations, in mind.

In this report, Turing characterised intelligence as requiring 'discipline', which he identified with the programmability of a universal machine, plus a residue of 'initiative.' Initiative now played the role that 'intuition' had done in 1938: mental actions apparently going beyond the scope of a 'definite method'. How was initiative to be found within the scope of computable operations, and so implemented on a computer?

Turing suggested various possibilities all based on imitating human brains: learning, teaching, training, searching. From the outset of his design work in 1945, Turing had been enthusiastic for exploiting the feature of a stored-program computer that a program can be manipulated in the same was as data. These ideas took his enthusiasm further, by having the machine actively modify its own programs. Turing emphasised that at a more fundamental level the concept of 'a machine changing its own instructions' was 'really a nonsensical form of phraseology', but it was convenient. The upshot of his argument was that by one means or another, and probably using many methods in combination, a computer could be made to imitate the mental functions of human brains.

From a purely biographical point of view, it is remarkable that someone so original, and whose individual qualities had generally been stoutly resisted by his social environment, should arrive at the conclusion that creativity is a computable process, something that could be achieved by a computer. But it was where he was led by his guiding idea of the brain as a finite machine, whose operations must be computable however different they appeared from what people had hitherto thought of as 'mechanical' in nature.

5. THE IMITATION OF MIND

This 1948 work was the background to the 1950 paper, in which Turing made a more public claim than ever before that intelligence can be evinced by computing machinery: i.e. belonged to the realm of computable processes. It was also a more ambitious claim than ever, since by provocative forays into the world of the Arts with witty talk of Shakespeare and sonnets, Turing made it quite clear that he was not restricting 'intelligence' to some special science-minded arena. The famous Test, pitting human against machine in demonstrating intelligence, embodied the 'fair play' announced in 1947. The setting of the test, however, with its remote text-based link, did have a further functional significance. It was supposed to give a way of separating those things Turing considered as relevant to intelligence, as opposed to other human faculties involving their many senses and physical actions. It is probably in drawing this distinction that Turing showed the least certainty, and this aspect of his paper has attracted the most criticism.

Returning, however, to Turing's central idea, it should be emphasised that Turing never imagined that the structure of the brain would resemble that of a computer, with a central control unit processing instructions serially. The crucial point here lies in Turing's exposition of the universal machine concept (p. 441). It follows from his argument that provided the operation of 'thought' is equivalent to the operation of *some* discrete state machine, it can be simulated by a program run on a single, universal machine, i.e. a computer. Thus Turing threw all his emphasis on the development of what would now be called software, rather than on the engineering of a brain-like object.

This point can be further refined. Turing's description of computability in the 1950 paper was all based on the finite capabilities of real, finite machines, illustrated by an account of the Manchester computer as it then stood. His claim was that the simulation of thought did not even require the full scope of computable functions, only that infinitesimal fraction of them which could be run using only a finite amount of 'tape'. (As a technical point, Turing's description did not even mention the 'tape'. This is because a finite tape can be absorbed into the instruction table of a Turing machine, which in turn he identified with the store of a computer such as the Manchester computer. This resulted in him rather confusingly describing the full gamut of computable processes as requiring an 'infinite store.' This is, however, just the unlimited supply of tape as prescribed in 1936, not an infinite instruction table.) He specified a necessary storage capacity of 10^9 bits, which of course is far surpassed by modern personal computers.

A fortiori, there is no suggestion in this paper of anything beyond the scope of computability. There were three areas of Turing's discussion where mathematics beyond the computable was raised, but in each case the thrust of Turing's argument was that computable operations would suffice. One of these was the Gödel argument, actually rather more weakly addressed in this than in his earlier papers, but still concluding that it had no force. The second lay in Turing's discussion of 'the continuity of the nervous system.' He claimed that the brain's basis in continuous matter, rather than being a discrete machine, was again no argument against the computability of thought: a discrete system could approximate a continuous one as closely as desired. The third was the concept of randomness, which Turing introduced without any serious definition. His illustration used 'the digits of π' as a random sequence, and this is *par excellence* a computable input.

In fact Turing's exposition also runs through two stages, reflecting what has been suggested above as his '1936' and '1941' stages of development. First the concept of computable functions, thought of as planned instructions (p. 438), and then followed the finite discrete state machine picture (p. 440). However, as he argued, these differently pictured machines could alike be implemented on a universal machine, the computer. This same two-part structure came into his final constructive proposals for the development of machine intelligence. Turing imagined rule-based programming (rather like expert systems as later developed), but also the 'child machine' learning for itself. Turing concluded by recommending that 'both approaches should be tried': he never suggested a rigid dichotomy between top-down and bottom-up approaches, which was later to divide Artificial Intelligence research so deeply.

In summary, Turing was able to claim:

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning.

This prophecy of the power of the computable was, of course, to stimulate the Loebner competitions as the dateline of 2000 approached.

6. AFTER THE TEST

No account of Alan Turing would be complete without a mention of his last years. *Computing Machinery and Intelligence* was shot through with courtroom images of juries and trials; they were prophetic. Turing was arrested for his affair with a young Manchester man in 1952. All homosexual behaviour was then illegal. He was seriously disturbed by the punishment that ensued: his brain was 'treated' with oestrogen. But his mind did not atrophy in this new period. In 1950 he had begun serious work, involving use of the Manchester computer, on a new theory of non-linear partial differential equations, proposed as 'the chemical basis of morphogenesis'. This and other research continued vigorously despite the interruption. (So did his personal life, which as usual he refused to adjust to the expectations of society.)

The question arises as to whether there were further developments in Turing's ideas about machine intelligence after 1950. There is an indication that there were. In the following year his popular talk on BBC radio (Turing 1951). It was basically a version of what he had set out in the 1950 paper. He explained the principle that any mechanical process could be simulated by a program run on a single machine, the computer: in particular, he had in mind the function of the brain. But this time he inserted an important *caveat* that had not been made in 1950. The machine to be simulated

> ...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.

Copeland (1999) has rightly signalled the importance of this new point, but his critical context suggests a link with the 'oracle,' a particular kind of uncomputable function used in Turing's 1938 work (Turing 1939). But Turing made no reference to this 'oracle' when admitting this possibly fatal flaw in his argument about the brain as a discrete state machine. The question he was raising was whether the space-time properties of quantum-mechanical physics could be captured by a discrete state machine model. And this was a question which went back to his earliest serious thought, being related to the work by Eddington and von Neumann that he was reading in 1928-32, especially that of Eddington.

In 1932 Turing had speculated, influenced both by Eddington, and by trauma in his personal life, that quantum mechanics underpinned free will (Hodges 1983, p. 63). The relationship between von Neumann and Turing has enjoyed much attention because of the question of who first had the idea of a practical universal machine (Hodges 1983, see also Davis 2000). Less well known is that Turing's first serious research study was of von Neumann's work on the foundations of quantum mechanics. Von Neumann clarified the measurement or *reduction process* in quantum mechanics; it is this which is not predictable. Seventy years later there is no agreed or compelling explanation of how or when such 'reduction' occurs. In 1953-4 Turing wrote to his friend and colleague Robin Gandy that he was 'trying find a new quantum mechanics' (Gandy 1954). Probably he was trying to make a predictable theory of the reduction process, to close the loophole in his argument for the computability of brain processes. However he died in 1954 before announcing any result.

His last published paper (Turing 1954), was again semi-popular, appearing in *Penguin Science News.* This did not mention quantum mechanics, but it went back to the pure mathematics of computability (which had taken recent new life with advances in algebra), and gave an account of Gödel's theorem. His conclusion was surprisingly unlike that pronounced 1950; he said that Gödel's theorem showed that 'common sense' was needed in interpreting axioms, and this time the intuitive 'common sense' was not asserted to be a something a machine could show as well as a human being. The year 1950 seems to have marked the peak of his dogmatism about machine intelligence, but it is impossible to know how his views would have developed had he lived longer.

In recent years Roger Penrose has taken up the two themes that Turing found most difficult to fit into his thesis of computable mental functions — Gödel's theorem and the quantum-mechanical reduction process — and has said that they must be connected (Penrose 1989, 1994, 1996). Penrose holds that the function of brain *cannot* be simulated by a computer program because of its quantum mechanical physical basis. Thus, for entirely materialist reasons, no truly intelligent behaviour will ever be simulated by a computer: the full Turing test will never be passed. Many commentators have attacked this conclusion, but it must be said that the topics Penrose emphasises are those that Turing himself found central to his arguments.

We are now so used to the properties of digital machines that is hard to imagine the world in which Turing conjured them from pure imagination in 1936. However it is crucial to see that what Turing offered in 1950, based on this earlier work, was something that went beyond the traditional mind-matter debate, and beyond loose science-fiction talk about humans and robots. It had a new solid substance in the digital or discrete-state machine model, made clear as never before. This structure, however, had a non-obvious limitation expressed by Gödel's theorem and Turing's own discoveries in the theory of computability. Turing always had these questions in mind. Turing's universal machine now seems to sweep all before it, and continues to captivate us with the apparently never-ending range of applications that it can encompass. Turing's own excitement for this project, his game-playing enthusiasm and iconoclastic humour, live on in every conversation-program writer of the present day. But it should be remembered that Turing's imitation game actually first arose as the 'fair play' argument for escaping the force of Gödel's theorem and the serious puzzle posed by the limits of what can be computed.