Cologne Summer School of Interdisciplinary Anthropology III
"The Phenomenality of Material Things:
Praxis – Genesis – Cognition"
18th to 23rd September 2017,
a.r.t.e.s. Graduate School for the Humanities Cologne

Alan Turing's models of cognition

Thomas Hainscho

18th September 2017

Abstract

The aim of this paper is to investigate Alan Turing's models of cognition and outlining the role of materiality in his understanding of cognition. I regard it as a contribution to the history of science but also take a systematic, philosophical standpoint with regards to the question of how material objects are involved in understanding cognitive processes. The paper has three parts; after a short introduction, the first part provides an overview of the orthodox reading of Turing as proponent of formal-logical understanding of the mind, the second part investigates the role of materiality within the paradigm of symbol based cognition and the third part contrasts the given findings with excerpts of Turing's writings on material aspects of machinery.

Introduction

There are several reasons for having heard of Alan Turing. As Andrew Hodges points out in his Turing biography, in 2011, former US president Barack Obama singled out Isaac Newton, Charles Darwin, and Alan Turing as British contributors to science (Hodges, 2014, p. xv). It is clear that Darwin stands for biology and evolution theory, and that Newton stands for physics, but what does Turing stand for? In informatics and mathematics, he is famous for his notion of the Turing machine and his theory of computability; in philosophy, there is the so-called Turing test, and the so-called Turing-mechanism is a concept in theoretical biology. Aside from these clearly academic contributions, Turing and a team of scientists successfully worked on a deciphering device for encrypted messages of the German National Socialists in the late 1930s. Despite the prominent connection between Turing and the Enigma, due to two mainstream films (*Enigma* from 2001 and *The Imitation Game* from 2014), his engagement in the secret service was unknown until several declassifications of secret service documents and publications in the 1970s (cf. Winterbotham, 1974). However, Turing can be regarded as a thinker, or scientist, with a strong interdisciplinary orientation – not simply because he contributed to different branches of science but because he transferred concepts from one discipline to another.

Despite his broad scientific engagement, Turing owns a certain reputation as mathematician and logician. This reputation directs the expectations regarding his view on cognition towards a certain direction. The general expectation is: what we can learn from Turing's conception of cognition is the formal-logical, hence symbol-based view. Positioning him as an interdisciplinary thinker, because of the concepts and methodical approaches from different disciplines, even confirms this view: He formed and endorsed a model of human cognition based on symbolic, computable processes and hence, counts as one of the key figures who propelled the conceptual change from the notion of machine intelligence (cf. Copeland, 2000, p. 519) to artificial intelligence (cf. Coppin, 2004, p. 9).

1 Turing machines and computability

The basis for the orthodox view on cognition is developed in Turing's text *On computable numbers, with an application to the Entscheidungsproblem* which was published in two parts in the years 1936 and 1937. It introduces the concept that will later be known as the Turing machine. This name was not coined by Turing himself but by Alonzo Church, who referred to the machine concept in a review of Turing's text (cf. Petzold, 2008, p. 63). In the following, I will give a basic overview of the Turing machine including a later text entitled *Intelligent Machinery* from 1948, which was posthumously published in 1969. Considering this later text allows for the claim, that Turing introduces two "varieties of machinery" (Turing, 1948, p. 3) in the 1936/37 text on computable numbers.

So, what is a Turing machine? Usually, a Turing machine is described as consisting of a tape and a device that operates on the tape. The tape is infinitely long and divided into discrete fields or cells. It is possible to write on the tape in the way that one cell contains one symbol of a predefined alphabet. In addition to the tape, a Turing machine has a device that can read, erase, and write the symbols of one cell on the tape, as well as shift to the next cell left or right hand.

A Turing machine runs in the following way: The starting setup consist of the tape carrying a certain input and the read/write device standing at first cell at the beginning of the tape. The Turing machine is in a predefined internal state S_1 that demands certain rules, like: Read the first symbol in the first cell c_1. If symbol X is read on the tape, move right to the next cell c_2 and change to state S_2; if symbol Y is read on the tape, move right to the cell c_2 and erase the content of the cell. Depending on the state and the scanned symbol, the machines performs certain actions, such as erasing or writing the symbol, or moving to another cell. Parsing in the inputs, moving, writing, and erasing, is all a Turing machine does. Finally, there are particular end states in which the Turing machine stops operating. This means that the computation is completed.

There is a formal notion of a Turing machine, so everything I just described can be written down with well-defined syntactic rules – a set of states with a subset of end states, a set of accepted input symbols, a transition function that maps a state and a symbol to a state etc. A typical task for a Turing machine is to parse in a certain number, which is written encoded on the tape, and write a symbol that depends on properties of the number. For instance, there is a Turing machine that decides whether a given input represents an even or an odd number, a Fibonacci number, or a member of a certain set. This may not sound overwhelming, but, in fact, these are problems or tasks of computers; and: despite its rather limited abilities, Turing provided formal proof that the concept of the Turing machine "[...] could be set up to churn out every provable assertion within Hilbert's formulation of mathematics" (Hodges, 2014, p. 133). This means that a Turing machine can solve every mathematical problem which has been shown as solvable. So far, there is no known problem with a solution that is restricted exclusively to human ingenuity. Vice versa, there are problems that cannot be solved by a Turing machine and these problems have not yet been solved by humans either. Delivering this proof was the paper's aim; Turing proofed that there is no solution to the Entscheidungsproblem – the decision problem –, which is addressed in the paper's title and refers to a mathematical problem stated by David Hilbert and Wilhelm Ackermann in 1928.

To deliver this proof, Turing introduced a formal concept that allows transforming the formal notation of a Turing machine into the input of a Turing machine's tape. This latter machine is the second variety of machinery that he introduced as "universal computing machine" (Turing, 1936/37, p. 241), and later on called universal Turing machine. This kind of machine always has a certain input, namely the formal standard description of a Turing machine. It does not execute one particular job, as the aforementioned Turing machine, but executes any job, depending on the input. In *Intelligent machinery* Turing summarises:

"The importance of the universal machine is clear. We do not need to have an infinity of different machines doing different jobs. A single one will suffice. The engineering problem of producing various machines for various jobs is replaced by the office work of 'programming' the universal machine to do these jobs." (Turing, 1948, p. 4).

So far, cognition has not yet played a role. The theoretical import of Turing's universal machine lies fully within the realm of mathematics. Its philosophical import lies in demonstrating the scope of his theory of computability. Turing claims, and this is a colloquial expression of the Church-Turing thesis, that the Turing machine's computability power equals the natural power of human computability. Unsurprisingly, the thesis has yet not been evaluated as right or wrong, since there is no definition of the natural power of human computability. Turing admits that "[a]ll arguments which can be given are bound to be, fundamentally, appeals to intuition, and for this reason rather unsatisfactory mathematically" (Turing, 1936/37, p. 249). Nevertheless, the thesis is generally expected to be true. The entry on Computability and Complexity in the *Stanford Encyclopedia of Philosophy* even reads: "[...] the 'Church-Turing Thesis', is uniformly accepted by mathematicians" (Immermann, 2015). Within the thesis, however, lies a little dirty trick. Finding formal proof by use of conventional mathematics assures that there is a Turing machine that can come to the same result.

One of Turing's "appeals to intuition" takes a surprising direction. As Hodges writes, the given pages rank "[...] among the most unusual ever offered in a mathematical paper [...]" (Hodges, 2014, p. 133).

> Computing is normally done by writing certain symbols on paper. We may suppose this paper is divided into squares like a child's arithmetic book. In elementary arithmetic the two-dimensional character of the paper is sometimes used. But such a use is always avoidable, and I think that it will be agreed that the two-dimensional character of paper is no essential of computation. I assume then that the computation is carried out on one-dimensional paper, i.e. on a tape divided into squares. [...] The behaviour of the computer at any moment is determined by the symbols which he is observing, and his "state of mind" at that moment. We may suppose that there is a bound B to the number of symbols or squares which the computer can observe at one moment. If he wishes to observe more, he must use successive observations. We will also suppose that the number of states of mind which need be taken into account is finite. [...] Let us imagine the operations performed by the computer to be split up into "simple operations" which are so elementary that it is not easy to imagine

> them further divided. Every such operation consists of some change of the physical system consisting of the computer and his tape. We know the state of the system if we know the sequence of symbols on the tape, which of these are observed by the computer (possibly with a special order), and the state of mind of the computer. (Turing, 1936/37, p. 249–250)

In this passage, Turing is talking about a computer but we may sincerely ask what does he mean by that. The first part of the paper was published in 1936, and consulting the *Oxford English Dictionary*, the first meaning of computer reads: "A person who makes calculations or computations; a calculator, a reckoner; *spec.* a person employed to make calculations in an observatory, in surveying, etc. Now chiefly *hist.*" (computer, n., 2017; cf. Hodges, 2014, p. 134). This means that Turing describes how a human calculates and step after step, the actions of the human calculator are transformed into the actions of a Turing machine.

So, the investigation of Turing's models for cognition has come to a result: The Turing machine delivers a mathematical theory of computability but also delivers a model of human cognition, precisely determining its computational power. I mentioned the classical description of the Turing machines as a tape and read/write device. However, the depiction of a Turing machine is open to imagination. Charles Petzold gives another picture and writes: "What does a Turing machine look like? You can certainly imagine some crazy looking machine, but a better approach is to look in a mirror" (Petzold, 2008, p. 71).[1]

2 Consequences for models of cognition

The fact that there is no known mathematical problem that could be solved by a human being and not by a machine, respectively, that all up-to-date human-solvable problems are equally machine-solvable, did give a boost to artificial intelligence research based on symbol manipulation. The mid-1940s analogy between

[1] Petzold's verdict seems to be unique for the recent years. Surprisingly, the idea of viewing Turing machines as human beings can be found in the writings of Ludwig Wittgenstein. Turing attended Wittgenstein's class on the foundation of mathematics in Cambridge (cf. Hodges, 2014, p. 193ff.). In §1096 of the *Bemerkungen über die Philosophie der Psychologie* (Remarks on the Philosophy of Psychology) Wittgenstein writes: „Turings 'Maschinen'. Diese Maschinen sind ja die *Menschen*, welche kalkulieren. Und man könnte, was er sagt, auch in Form von *Spielen* ausdrücken. Und zwar wären die interessanten Spiele solche, bei denen man gewissen Regeln gemäß zu unsinnigen Anweisungen gelangt. [...]" (Wittgenstein, 1984, p. 197; RPP I, §1096); – "Turing's 'Machines'. These machines are *humans* who calculate. And one might express what he says also in the form of *games*. And the interesting games would be such as brought one via certain rules to nonsensical instructions. [...]" (Wittgenstein, 1980, p. 191).

the brain and electronic computers became the dominant paradigm in both, the 20$^{\text{th}}$ century neurosciences and artificial intelligence research. As philosopher of science Peter Asaro (2011) points out, "[...] the computer was from its very conception a kind of model of the brain. Researchers in Artificial Intelligence or Cognitive Science did not 'discover' any analogy between the mind and the computer." (Asaro, 2011, p. 103). Hence, the brain/computer paradigm developed reciprocally. The computer was modelled after views in neurology and, vice versa, terminology and guiding principles in neurosciences were lent from computer science.

There are historic sources for this claim. Prior to the 1940s, electronic calculator were constructed for one particular task, like the *Rockefeller Differential Analyzer* (1942), which solved differential equations – and could only solve differential equations. Other multi-purpose machines had to be physically changed ("rewired"), in order to carry out new tasks. The general-purpose computer was different; all software was able to perform on one hardware. It appears to be a construction of the Universal Turing machine. Klara von Neumann describes the interlink between computer and brain science in her husband's, John von Neumann, work:

> After the war, together with a small group of selected engineers and mathematicians, Johnny built, at the Institute for Advanced Study, an experimental electronic calculator, popularly known as JONIAC, which eventually became the pilot model for similar machines all over the country. Some of the basic principles developed in the JONIAC are used even today in the fastest and most modern calculators. To design the machine, Johnny and his co-workers tried to imitate some of the known operations of the live brain. This is the aspect which led him to study neurology, to seek out men in the fields of neurology and psychiatry, to attend meetings on these subjects, and, eventually to give lectures to groups on the possibilities of copying an extremely simplified model of the living brain for man-made machines. (von Neumann, 1958, p. viii)

What impact does this paradigm have on the conception of cognition? Materiality does not play an important role. The brain/computer-paradigm supports the view that the human mind is a symbol-processing power that somehow happens to be materialised in the brain, but it could as well be materialised in a computer. Cognition can be simulated by a material device, at least in terms of computation. Turing wrote about the pen and paper simulation (cf. Turing, 1936/37, p. 249–252; Turing, 1948, p. 5), but any sufficient material will do;

Joseph Weizenbaum (1976, p. 51ff.), for instance, ironically suggested to construct a Turing machine with a roll of toilet paper. In comparison, the human mind and all material realisations are more restricted than the Turing machine because it, as a model, operates on an infinite tape with unlimited time for its calculations and independent from physics.

Reading Turing's 1950 paper *Computing Machinery and Intelligence*, which introduces the Imitation Game, allows assuming, that Turing himself did not care about materiality either. The Imitation game, to which is informally referred to as the "Turing test", follows a rather pragmatic approach concerning cognition. Instead of investigating the question "Can machines think?", Turing suggests to play the Imitation Game, that is about deciding, whether a written answer comes from a machine or a human being. Here again, all material aspects are left out of the game.

The conception of the mind results in something that is instantiated in matter but works on logical principles, independently from physics. The mind has certain mental states, and being in a certain state leads to a certain action. As an entity that can carry out an infinite number of tasks, the mind is basically a static concept and neither able to develop or even to change. This circumstance was criticised by Kurt Gödel (cf. Petzold, 2008, p. 192).

3 Turing's account on materiality

Despite being the common approach to modelling the mind, symbolic accounts have been criticised in recent years. Theories like the thesis of Radical Embodied Cognition (cf. Clark, 1997) or enactivism (cf. Varela, Thompson & Rosch, 1991) offer alternatives and take the role of materiality into account. A. J. Wells (1998), for instance, brings the Turing machine's tape together with extended memory and environment intelligence. But I am not interested in favouring one position over another but in historically searching for Turing's account on materiality.

The Turing machine is a mathematical model and materiality does not play a relevant role for it. However, Turing's biography offers some interesting insights concerning material aspects, since it appears that he was interested in material aspects of computation. Hodges provides historic evidence, that Turing understood his own work as the construction of a brain (cf. Hodges, 2014, p. 364f.). Elizabeth Wilson characterises him as being just as interested in the theory of computational logic as in building mechanical and electronic devices. During his time in Bletchley Park, when he worked for the British Secret Service in the late 1930s, he "spent as much time with wires and valves and reverie as he did with mathematical calculation" (Wilson, 2010, p. 39).

From 1945 to 1946, he worked at the National Physics Laboratory in London, where he was involved in planning and constructing the Automatic Calculating Engine, an early digital computer (cf. Wilson, 2010, S. 32–33). He contacted psychiatrist and cybernetic pioneer W. Ross Ashby and expressed his interest in collaboration: "In working on the ACE I am more interested in the possibility of producing models of the action of the brain than in the practical applications to computing." (Turing, 1946). In 1947, Turing left the NLP and the director of the institution, Sir Charles Darwin, wrote in a letter to Sir Edward Appleton from 23$^{\mathrm{rd}}$ July about Turing's going off:

> He wants to extend his work on the machine still further towards the biological side. I can best describe it by saying that hitherto the machine has been planned for work equivalent to that of the lower parts of the brain, and he wants to see how much a machine can do for the higher ones; for example, could a machine be made that could learn by experience? (Darwin, 1947)

Apart from the biographical indications that Turing did in fact consider materiality in modelling cognition, there are some findings in his texts:

- A common reference is made in respect to materiality as a flaw. A material machine wears off and needs to be repaired from time to time. If a machine is suited to learn and to be educated by a "competent schoolmaster" (Turing, 1996, p. 256), it may also need a mechanic who repairs it as he writes in the unpublished text *Intelligent Machinery, A Heretical Theory.*

- In the same text he contrasts the desires of a machine with human desires. He mentions material desires and for him, there are machine equivalents: "Let us suppose that [...] owing to its having no hands or feet, and not needing to eat, not desiring to smoke, it will occupy its time mostly in playing games such as Chess and GO, and possibly Bridge" (Turing, 1996, p. 257). This contrast is more than a funny side remark. Turing seriously tries to state machine equivalents for human needs (cf. Turing 1948, p. 9) as well as human abilities like sex or sports. He sketched a machine creativity with a component for random number generation (cf. Turing, 1950, p. 451).

- In different texts (Turing 1948; 1996) he wrote about the analogy between man and machine in a broader sense and described which material devices could provide proper input for a computer, like keyboards, video cameras and microphones etc. as analogies for sensory organs.

Finally, another finding in Turing's texts is given in a text about the concept of hypercomputation Jack Copeland and Diane Proudfoot (cf. Copeland & Proutfood, 1999). In 1948's *Intelligent Machinery* report, Turing introduced the concept of the unorganized machine. An unorganised machine is made up of a large number of similar units. These units are very simple and possess two input channels and one output channel. A large number of these units is connected with each other, forming a network. Aa signal runs through the input channel and Turing describes that and how the network of interconnected units justifies itself for certain signal flows (cf. Turing, 1948, p. 6–7, 11–13; Copeland & Proudfoot, 1999, p. 100). Since the report remained unpublished and Turing's remarks are rather short, the concept remained rather unknown. As Copeland notes, Turing's unorganised machines provide a very early example of "networks of neuron-like elements" (Copeland, 2004, p. 403) and Turing himself describes it as "the simplest model of a nervous system" (Turing, 1948, p. 6). Turing then continues to compare it to the human brain: "All of this suggests that the cortex of the infant is an unorganised machine, which can be organised by suitable interfering training" (Turing, 1948, p. 12)

What are the consequences of relating the unorganised machines network to an infant's brain cortex for models of cognition? It appears that Turing did in fact consider materiality in his models, since the unorganised machine network is intended to function as artificial brain – so, there is materiality.

Conclusion

I think – and these are my concluding remarks – that there are two aspects of cognition modelling to be found in Turing. The first aspect are the models for cognition on a symbolic logical level. Turing claimed that the Turing machine provides an accurate model for human cognition, and, being a model, exhibits similarities with and differences from its target system. The human mind cannot operate with infinite time or infinite memory etc. Materiality limits computability, it is responsible for flaws but contributes to humanising the machine model by increasing the similarities to the human mind. Hence, the assumption that Turing believed his work to be building artificial brains with a natural cognitive power has a solid theoretical ground.

The second aspect for cognition can be found in Turing's anticipatory views on neural networks. Although I do not share the conclusions given by Copeland & Proutfoot (1999) about hypercomputation, I agree that there is a difference between the concept of the unorganised machines as "model of the cortex" (Copeland & Proutfoot, 1999, p. 100) and the Turing machine. However, by claiming that an infant's brain is an unorganised machine network, Turing did not state a model, he stated identity. Speaking of interconnected input/output units is speaking of neurons. Hence, we do not deal with a possible representation of cognition but with cognition itself, which is not materialised in a machine but as a machine. Of course, cognition in a certain manner. Undoubtedly, Turing is a materialist at heart; he is even a reductive materialist and believes that the physical equipment is sufficient to produce all mental phenomena. Nevertheless, he does not believe that the human brain literally is a Turing machine. The claim that the brain is an unorganised machine network can be found in his writings. This provides a material account on cognition, which is not a model but a scientific explanation in theoretical terms.

References

Clark, A. (1997). *Being There. Putting Brain, Body, and World Together Again.* Cambridge: MIT Press.

computer, n. (2017). *Oxford English Dictionary Online* (June 2017). Online: http://www.oed.com/view/Entry/37975

Copeland, B. J. & Proudfoot, D. (1999). Alan Turing's Forgotten Ideas in Computer Science. *Scientific American*, (280), 98–103.

Copeland, B. J. (2000). The Turing Test. *Minds and Machines*, *10*(4), 519–539.

Copeland, B. J. (2004). Introduction to Intelligent Machinery (1948). In B. J. Copeland (Ed.), *The Essential Turing* (pp. 395–409). Oxford: Oxford University Press.

Coppin, B. (2004). *Artificial Intelligence Illuminated.* Sudbury: Jones & Bartlett.

Darwin, C. (1947). *Letter to Sir Edward Appleton* (23 July 1947). The Turing Archive for the History of Computing. Online: http://www.alanturing.net/darwin_appleton_23jul47/

Hodges, A. (2014). *Alan Turing: The Enigma.* London: Vintage.

Immerman, N. (2015). Computability and Complexity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition).* Online: https://plato.stanford.edu/archives/spr2016/entries/computability/.

Petzold, C. (2008). *The annotated Turing. A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine.* Wiley Publishing: Indianapolis.

Turing, A. M. (1936/37). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*(3/4), 230–265.

Turing, A. M. (1946). *Letter to W. Ross Ashby* (approximately 19 November 1946). The W. Ross Ashby Digital Archive. Online: http://www.rossashby.info/letters/turing.html

Turing, A. M. (1948). *Intelligent Machinery.* Unpublished report. National Physical Laboratory. Online: http://www.alanturing.net/intelligent_machinery/

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460.

Turing, A. M. (1996). Intelligent Machinery, A Heretical Theory. *Philosophia Mathematica*, *4*(3), 256–260.

Varela, F., Thompson, E. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge: MIT Press.

von Neumann, K. (1958). Preface. In John von Neumann, *The Computer and the Brain* (pp. v–x). New Haven: Yale University Press.

Wells, A. J. (1998). Turing's Analysis of Computation and Theories of Cognitive Architecture. *Cognitive Science*, *22*(3), 269–294.

Weizenbaum, J. (1976). *Computer Power and Human Reason.* New York: W. H. Freeman & Company.

Wilson, E. A. (2010). *Affect and artificial intelligence.* Seattle: University of Washington Press.

Winterbotham, F. W. (1974). *The Ultra Secret.* London: Weidenfeld and Nicolson.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology.* Vol. 1. Edited by G. E. M. Anscombe, and G. H. von Wright, translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, L. (1984). Bemerkungen über die Philosophie der Psychologie. In L. Wittgenstein, *Bemerkungen über die Philosophie der Psychologie.* Werkausgabe Bd. 7 (S. 7–346). Frankfurt am Main: Suhrkamp.