February 2009, minor revision in 2014
University of The Arts Bremen
Digital Media Master's programme
Frieder Nake – *Ars Algorithmica*

Andre Hoffmann – contact@effzehn.de

Why humans might think that machines might (not) think

The Big Question: Can machines think?

In times machines were considered to be just mechanical tools, not more sophisticated than a hammer or a lever, the question if a machine can think would be absurd. But as we know, some machines developed from primitive tools to highly complex systems, allowing mankind to perform complex calculations in practically no time, travel with high speeds at a relatively high security level (cars and airplanes), let us dive deep into virtual realities or allowing us the flight to the moon. It is this steady development which inevitably led to the question what else machines might be able to do.

In 1950, Alan Turing's paper "Computing Machinery and Intelligence" was published in the Journal "Mind". He begins his work with the following words: "I propose to consider the question, ‚Can machines think?'" [Turing, 1950, p.50] Since Turing himself stated, that he found the original question „too meaningless to deserve discussion" [Turing, 1950, p.55], he approached the answer from an empirical perspective, a perspective which circumstances he called the "Imitation Game". We nowadays know it under the name *Turing-Test.*

Turing was not the first to consider those types of questions. In fact, Descartes already discussed the distinguishability of humans and machines in the 17th century. [Shieber, 2004, p.17] He was "perhaps the first real champion of the mechanization of mind." [Husbands et. al, 2008, p.4]

* * *

We will address this and some more writings in direct connection with Turing‘s paper. Turing stated that "We cannot altogether abandon the original form of the problem, for opinions will differ as to the appropriateness of the substitution and we must at least listen to what has been said in this connection." [Turing, 1950, p.55]
So he argued some contrary views in the same paper, considering some thoughts made before his time.

* * *

The Turing-Test

To follow the discussion, it is important to understand the setup of the Turing-Test.
Turing‘s original setup for this game consists of three persons: a man, a woman and an interrogator who may be of either sex and stays in a room apart from the other two. The task of the interrogator is to determine which of the other two is the man and which is the woman. He knows the two by labels, e.g. X and Y and states his assessment in the end by saying X is A and Y is B or X is B and Y is A.
The answers given to the interrogators questions should be typewritten or could also be repeated by an intermediary. It is important that there is no direct connection between the two and the interrogator for that the interrogator cannot get any more clues.
Turing’s eminent question is now: What will happen when a machine takes the part of one of the players in this game? Will the interrogator make the same decisions as when the game is played with a man and a woman?
The Turing-Test nowadays is generally formulated a little bit more uncomplicated: the question is, if the interrogator can tell if X and/or Y is a person or a computer by stating questions in natural language.

* * *

Turing thought that the imitation game may lead to an answer to his original question “Can machines think?” He later describes that with machines he means digital computers to exclude every other kind of possible or thinkable workarounds which might be similar to the concept of a machine. But he saw the technical boundaries to realize such a machine since of the lack of memory capacity and processing speed. Having explained in his paper what a digital computer is, he refines his rather unsuitable and placative question to “Are there imaginable computers which would do well in the imitation game?”

Turing believed that by the year 2000 computers would be so powerful that an "average interrogator would not have more than a 70% chance of making the right identification after five minutes of questioning". This prognosis is rather vague, not making us able to reliably answer it, nevertheless, until today there is no program which is able to pass the modern Turing-Test. [Weber, 2008]

* * *

Contrary Views on the main question

To discuss on his main question, Turing himself considered opinions opposed to his own. Doing so, he possibly wanted to scotch any upcoming discussion (see Dennet's statement below). Nevertheless, his views were highly argued. We can use his enumeration of objections to connect the discussion directly with his dramaturgy.

* * *

The theological objection: 'God has only given men and women the ability to think, hence no machine or animal can think.' But since God is supposed to be almighty why would he not be able to give machines or animals souls „if He sees fit"? [Turing, 1950, p.55]
Descartes already saw an analogy between machines and humans. In fact, for him machines, animals and human had the same functionality, comprised in the body of the subject. But he had the opinion that a human could never create a „machine" which is of the same quality as a human being, i.e. endowed with an immaterial and immortal soul and a mind, since this would be only Gods capability. Animals would have only a soul which is mortal and *not* endowed with a mind. His distinction of body and mind is known as substantial dualism. In *Discourse on the Method* he writes: „For they [engineers or the like] will regard this body as a machine which, having been made by the hands of God, is incomparably better ordered than any machine that can be devised by man, and contains in itself movements more wonderful than those in any such machine." [Shieber, 2004, p.27] Descartes also asserted that „there are no men so dullwitted or stupid – and this includes even madmen – that they are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal, however perfect and well-endowed it may be, that can do the like." [Shieber, 2004, p.28] This seems to be a basic anticipation for the Turing-Test. However, from a systematic point of view Descartes opinions were inept. Although his reasoning is thought well over his perspective stands on a very vague fundament. Dietz described the lack of his argumentation: „Descartes' substance dualism is full of

inconsistencies. It is unexplained if and how body and mind affect each other. The elsewhere so lucid rationalist only makes vague indications." [Dietz, 2003, p.106] Back to Turing's argument, he himself states that he is „not very impressed with theological arguments whatever they may be used to support. Such arguments have often been found unsatisfactory in the past." [Turing, 1950, p.55]

* * *

The heads in the sand objection: 'The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so.' Turing writes that this argument is in a way connected to the religious argument and is founded on the believe of the superiority of humans, being the last instance between a deity and the world. He does not think this argument is substantial enough to refute it. Shieber rules that this argument is „frivolous". [Shieber, 2004, p.63] In D.C. Dennetts discussion, the questioner asks why it does „seem that some people are upset by AI research? Does AI threaten our-self esteem?" Dennett introduces Herb Simons explanation to this phenomena: „For many people the mind is the last refuge of mystery against the encroaching spread of science, and they don't like the idea of science engulfing the last bit of *terra incognita.* [...] This could lead to the ‚evil scientist' who can control you because he or she has a deep understanding of what's going on in your mind. This seems to me to be a totally valueless fear, one that you can set aside, for the simple reason that the human mind is full of an extraordinary amount of detailed knowledge [...]." [Shieber, 2004, p.289]

* * *

The mathematical objection: ‚Not everything is computable and so may be the human thought.' Kurt Gödel shocked the world of mathematics by proving that the field of mathematics itself must be incomplete since it is not possible to calculate (or: compute) certain phenomena, e.g. it is not possible to compute if a Turing machine, normally configured with a ruleset and an alphabet, will stop once it's switched on or goes on forever. This problem is known under the name *Halting Problem.* Turing saw a parallel between the phenomena of thought and Gödel's so-called *Entscheidungsproblem*. If we convert the Halting problem into textual speech the question we might ask in the Turing-Test would be: „Consider the machine specified as follows.… Will this machine ever answer ‚Yes' to any question?" [Turing, 1950, p.56] According to the subject of the Halting Problem, the machine would never be able to answer the question correctly. In

contrast, human beings are considered being without any of „such limitations“, although there is no „sort of [any] proof.“ A machine which would give any definite answer to this critical question must be wrong, which might give the interrogator a feeling of „superiority“. Nevertheless, there „might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on.“ In fact, we are observing this race between machines and human in the game of chess, where computers beat humans and humans beat computers. Turing himself somehow foresaw this. In *Proposed Electronic Calculator* he writes: „‘Can machines play chess?‘ [...] 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.“ [Teuscher, 2004, p.331]

* * *

The argument from consciousness: ‚If intelligence and conscience are bounded together, the proof of consciousness would be the proof of intelligence at the same time.‘ Dennett states: „I do think that it‘s possible to program self-consciousness into a computer. Self-consciousness can mean many things. If you take the simplest, crudest notion of self-consciousness, I suppose that would be the sort of self-consciousness that a lobster has: when it‘s hungry, it eats something, but it never eats itself.“ [Shieber, 2004, p.290] Here the question remains if consciousness is an observable property of a subject or if it is an inward-only quality. Turing introduced a solipsistic point of view. One cannot ever be sure if another being is really conscious like oneself. We only assume that a person or an animal is conscious. The only way to find out how a human thinks (is conscious) is to be that very human. Dietz comments on the ongoing debate of consciousness: „[...] but in fact there is nothing else behind it than a remake of the well-know Body-Mind problem.“ [Dietz, 2003, p.261]

* * *

Arguments from various disabilities: ‚There is a lot more humans can do.‘ Turing addresses some arguments which are seemingly not applicable to machines, like “have a sense of humor”, “enjoy strawberries with cream”, “fall in love” or “do something really new”. Turing states that those human-specific abilities are either non practical, could be realized with enough storage capability or could be easily imitated. The question is: for what reason should those abilities ever be imitated? Schnelle writes: „Studying, or even trying to implement, co-textual and contextual background evaluation in humans may appear

to be silly as long as our perspectives are determined by definite and circum-scribed designs. However, it is a challenge to come up with a better understanding of that powerful ability applied in solving the many practical tasks we are confronted with in our human affairs. This is also a property of human thought – and not of the „thoughts" of machines – a property whose machine implementation Turing correctly characterized as idiotic." [Teuscher, 2004, p.357]

* * *

Lady Lovelace's objection: Charles Babbage described the so-called *Analytical Machine* for the first time in 1837. This machine was a mechanical calculator which initially only existed as a draft. His colleague, the British mathematician Lady Lovelace (daughter of the poet Lord Byron) was heavily influences in Babbage's work. She added many notes to Babbages paper in which the Analytical Machine was described. One note illustrates her open mindedness by stating that the machine could act as a „thinking, reasoning machine" [Husbands, 2006, p.1] Despite this heavily optimistic and futuristic conclusion she interjects that „the Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform" [Turing, 1950, p.59], meaning „to say that the machine can *only* do what we know how to order it to perform." [Shieber, 2004, p.115] To refute this argument Turing introduces the concept of learning machines (which he thoroughly deals with in the last part of his paper). Purtill states that „hardly any behavior of thinking beings can be shown to be completely determined by ‚programming' (i.e. teaching or conditioning) plus external ‚inputs' (i.e. experience of various kinds)" whereas „any computer output can be explained along these [program] lines". [Shieber, 2004, p.170] In fact, learning is a critical factor if the aim is to design an intelligent machine. „If we ever do make an intelligent machine, presumably we will do it by equipping it with mechanisms for learning, solving problems, etc." [Shieber, 2004, p.248]

* * *

Argument from Continuity in the Nervous system: Turing was possibly the first person „to consider building artificial computing machines out of simple, neuron-like elements connected together into networks in a largely random manner." [Teuscher, 2004, p.334] So far, he must had some knowledge in this area, additional to his experience with discrete state machines. Therefore he was able to compare the nervous system with the discrete state machine, the digital computer. In this argument he was considering that a

discrete state machine cannot mimic a nervous system since it is not a nervous system with its distinct capabilities and complexities. The question is not, how an intelligent system is composed but rather what the output of this system might create.

* * *

The Argument from Informality of Behavior: Turing objected that it is not possible to formulate a set of rules to manage a persons everyday life and to react on every thinkable situation. Since discrete state machines rely on rulesets, transfered to the human it would mean: "If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine." [Turing, 1950, p.60] To the contrary, if human's behavior would rely on a ruleset, a machine could be constructed by this ruleset and therefore programmed to think as a human. But there is not such a ruleset found yet. Nevertheless, we never can say for sure, that there is actually not such a ruleset only because we have not found it yet.

It is particular interesting in this case that Turing obviously does not take into account his conclusions from Lady Lovelace's objection, the fact that humans learn and gain experience. Learning is some sort of the act of constituting a ruleset, though dynamically changing evermore. If we are able to program learning behaviors in machines, the presence of such a ruleset would be rather irrelevant for finding an answer to the main question.

* * *

The Argument from Extra-Sensory Perception: Turing even addresses the ideas of telepathy, clairvoyance, precognition and psycho-kinesis, together known as E.S.P. It is certainly quite frivolous to take those ideas into account since those phenomena leave the field of serious science. But nevertheless this argument reminds us to think about phenomena that might not be discovered yet but still have an effect on our life. „The idea that our bodies move simply according to the known laws of physics, together with some others not yet discovered but somewhat similar, would be one of the first to go." [Turing, 1950, p.60]

* * *

More views and discussion on the Big Question

The Turing Test is one highly regarded setup in the field of intelligence science. Turings original setup was criticized, reinterpreted (e.g. the inverse Turing-Test, where the

interrogator is a computer), amended, and inspired a lot of different thought experiments and certainly a specific terminology.

* * *

One critique is that the determination of intelligence is, in Turing's case, exclusively connected to the use of language. The language-dependent test was criticized as an insufficient approach for ascertaining if something or someone is intelligent. P.H. Millar writes: „[...] to put a five-year-old white Canadian and a seventy-year-old Pigmy into the Imitation Game test in order to decide a hypothesis that one or the other of them was not intelligent would be ludicrous." [Shieber, 2004, p.178] In contrast to that, Robert M. French states that the Turing Test, „provides a sufficient condition for human intelligence but does not address the more important issue of intelligence in general. [...] I feel that any attempt to „fix" the Turing Test so that it could test for intelligence in general and not just human intelligence is doomed to failure because of the completely interwoven and interdependent nature of the human physical, subcognitive, and cognitive levels. To gain insight into intelligence, we will be forced to consider it in the more elusive terms of the ability to categorize, to generalize, to make analogies, to learn, and so on." [Shieber, 2004, p.196 et seq.] One specific thought experiment deals with this very human-specific properties and abilities of the human mind. It is called the Chinese Room experiment.

* * *

It was in 1980, when the philosopher John Searle formulated this thought experiment in his paper "Minds, brains, and programs". He tried to disprove that the human intelligence can be simulated by computer programs by constituting the Chinese Room thought experiment. Searle summarizes his experiment the following way:

"Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese." [Wilson, Keil, 2001, p.115]

Searle wanted to point out that a person, who acts like a computer and manipulates symbols and has no connection to those symbols, he cannot understand those symbols or, more importantly, learn from those. The simple use of symbols is not sufficient for abilities like insight, perception, understanding or thinking. And since computers are machines for manipulating symbols, the execution of a program does not fulfill the requirements of an intellectual activity.

Searle's experiment created a debate and raised further discussions which made it evident that the meaning of intelligence or understanding can be viewed from different perspectives.

One can interpret the room as a combination of an executing part (the human), data and instruction. From a functional point of view, this is already a „thinking" system.

Another critique is doubts a possible validation of the experiment. There is no word on the length of his ruleset. If it will be found out that the ruleset in itself cannot exist (e.g. because is must be infinitely long) the experiment as a whole would be irrelevant.

* * *

Turing's question, the „Big Question", is so highly discussed that one might get the impression every word in the paper was interpreted and analyzed. The critical question is if the interpretation went the right way. Searle writes: „The Turing test is typical of the tradition in being unashamedly behavioristic and operationalistic [see argument (7)], and I believe that if AI workers totally repudiated behaviorism and operationalism much of the confusion between simulation and duplication would be eliminated." [Shieber, 2004, p.221]

Some involved stated that Turing's original aim behind writing his paper was misinterpreted while his original aim, actually to smother any upcoming discussion, was certainly not achieved. Dennett writes in his essay *Can machines think?*: „It is a sad irony that Turing's proposal has had the exact opposite effect on the discussion of that which he intended. Turing didn't design the test as a useful tool in scientific psychology, a method of confirming or disconfirming scientific psychology, a method of confirming or disconfirming scientific theories or evaluating particular models of mental function; he designed it to be nothing more than a philosophical conversation-stopper." [Shieber, 2004, p.270]

* * *

Also, the question if machines can have intelligence made different disciplines to appear on scene. Shieber suggests that the question rather should be „whether machines can imitate human intelligence“ to separate this question from the debate „on the applicability of anthropomorphic terms to non-humans.“ [Shieber, 2004, p.179]

* * *

Hubert L. Dreyfus, an opponent of the "artificial intelligentsia" [1] of the 60s, based his arguments on Heideggers‘ phenomenology. In his critique he opposes the fixation of rules of machines (in algorithmic form) to the being-at-hand of an external world for human beings – the context, experienced in the human body. [Dreyfus, 1979, p.235 et seq.] According to him, a big part of human knowledge is deeply interconnected within the structures of human action and the context of this action and not representable algorithmically inside a machine. He describes this non-representable knowledge as the „knowing-how“, according to that the representable (through symbols) knowledge as „knowing-that“. [KI Kritik]

* * *

Critique on the question

The question if a machine could pass the Turing test might be easier to answer than the question Can machines think? since we now know that language must not necessarily be the indicator for intelligence but only *one* indicator. The discourse amongst a variety of disciplines like psychology, philosophy, computer science, neuro-science, and others will certainly go on, since every discipline uses it‘s own models and explanations. If the notion of intelligence will be ever able to be summed down to a formula (which would be a critical condition for consciously constructing an intelligent machine) remains highly questionable. Jürgen Habermas states: „The ontologization of insight in natural science towards a naturalistic, on hard facts founding, shrinked view of the world is not science but bad metaphysics.“ [Dietz, 2003, p.264]

Time will tell if the organizations around the globe, struggling to construct a program which is able to pass the test, will fulfill their own foresights to pass the Turing test in the coming years, but others had been ambitious before and failed blatantly. [Weizenbaum, 1978, p.187] [2]

From a contemporary point of view, the question if machines can think is blurry and probably might never be answered since it is not clear what thinking actually means. There are of course many different approaches in handling the notion of thinking and, as stated before, those notions come from various scientific and philosophical backgrounds. However, the counter question if machines can think is not „what does thinking mean" but why do one ask this question anyway? As it is often the case this question may be so attractive because it will tell us something about ourself, in this case because the test „reveals a connection between possible computer activities and our ordinary concept of thinking" [Shieber, 2004, p.298] Turing writes: „The whole thinking process is still rather mysterious to us, but I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves." [Shieber, 2004, p.116]

* * *

Doing research in AI is a delicate task, facing the complexity while making very small achievements to serve humans in very specific and specialized ways. Alan Turing knew that *Can machines think* is a question too big too answer, a question which asks for everything and nothing, a metaphysical question, disregarding the fact that we live in a complex world and that we cannot deny the complexity in order to remotely understand thinking.[3] Like Searle pointed out, thinking, for human standards, needs consciousness, reason, knowledge, feelings, experience, foresight, identity, creativity, furthermore having this very own physical body, having neurons and a chemo-electrical nervous system and probably more abilities and properties we do not know yet. And the Turing Machine, a simple construct for symbolic computation, is supposably light years afar from this complexity. Turing, of course, was aware of this. His proposal may be seen in the light of a distinct frivolous british humor and to scotch upcoming discussions.

* * *

In particular the AI research underestimated the complexity of the topic and tried to imitate something they do not even fully understand, in fact not even slightly, when we look at the AI discourse of the 60s, the ELIZA case, a prove for the arrogance and naivety of the „artificial intelligentsia". From that point of view, the first mistake people in the popular AI science have committed is setting the standard for the notion *thinking* on a very low level, leading to a sell-out of the notion on the shelfs of electronic retailers. Intelligence became something everyone can buy and nothing special which is worth to strive for, but machines can substitute. The second mistake is that their analyses and

simplifications of the human mind, together with the universal claim of the scientific impeccability created the ghosts they called for: the mechanized minds – people believed that they were nothing but an automaton actually tended to behave like that. Science nowadays has the power to form us and our self-understanding although there were many revolutions and paradigms in science which had proven that science is sometimes not much more than just temporary beliefs, especially regarding the theme question.
Fortunately, most modern AI research is not interested in the question *Can machines think* anymore. It became more interesting to focus on certain parts of what constitutes intelligence and thinking. It is important to clarify the state of the art and, facing the facts of the failures in the ambitious AI research, it is time to award the vital quality of the term back to a universal but significant property of life as we know it.

NOTES

[1] a term coined by the probably most prominent AI critic, Joseph Weizenbaum

[2] more recently: [BBC NEWS]

[3] In the context of the theme complexity I recommend D. Mitchell, Sandra *Biological Complexity and Integrative Pluralism*. Cambridge, Massachusetts: Cambridge University Press, 2003

LIST OF REFERENCES

[Shieber, 2004] Shieber, Stuart (Ed.). *The Turing Test. Verbal Behaviour as the Hallmark of Intelligence.* Cambridge, Massachusetts: MIT Press, 2004.
[Teuscher, 2004] Teuscher, Christof (Ed.). *Alan Turing: Life and Legacy of a Great Thinker.* Berlin: Springer Verlag, 2004.
[Husbands et. Al, 2008] Husbands, Holland, Wheeler (Eds.): *The Mechanical Mind in History.* Cambridge, Massachusetts: MIT Press 2008.

[Turing, 1950] Turing, Allan. *Computing Machinery and Intelligence,* originally publicized in Mind: A Quarterly Review of Psychology and Philosophy 59(236): 433-460. October 1950. Reprinted in: Wardrip-Fruin, Montfort (Eds.). *The New Media Reader.* Cambridge, Massachusetts: MIT Press, 2003. Page numbers indicate the pages in this reprint.

[Dietz, 2003] Dietz, Peter. *Menschengleiche Maschinen. Wahn und Wirklichkeit der künstlichen Intelligenz.* Berlin: Bühler & Heckel, 2003.

[Weizenbaum, 1978] Weizenbaum, Joseph. *Die Macht der Computer und die Ohnmacht der Vernunft.* Frankfurt am Main: Suhrkamp Taschenbuch Verlag, 1978.

[Dreyfus, 1979] Dreyfus, Hubert L. *What Computer Still Can't Do*. Cambridge, Massachusetts: MIT Press 1979.

[Wilson, Keil, 2001] Wilson, R.A. and F. Keil (eds.). *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge, Massachusetts: MIT Press, 2001

[Weber, 2008] Weber, Jutta. *Nackte Maschinen.* c't Magazin für Computertechnik 15/2008. Hannover: Heise Verlag, 2008.

[Husbands, 2006] P. Husbands. Machine Intelligence, In. A.C. Grayling and A. Pyle and N. Goulder (Eds), *Continuum Encyclopedia of British Philosophy*, Thoemmes Continuum Press, 2006.
<http://www.cogs.susx.ac.uk/users/philh/pubs/Machine%20IntelligenceShort.pdf>

[BBC NEWS] *In conversation with... a computer program.* 16 October 2008. BBC. 9 November 2008. <http://news.bbc.co.uk/2/hi/uk_news/magazine/7670050.stm>

[KI Kritik] Hoffmann, Achim, *Die phänomenologische Kritik an der KI*. 12 July 2002. UNSW CSE. 19 January 2009.
<http://www.cse.unsw.edu.au/~achim/Research/Philosophie/node62.html>

Note to [Dietz, 2003] and [Weizenbaum, 1978]: Translations from German into English by the author