doi: 10.14355/ijps.2016.04.004

# Turing's Error-revised

Pertti SAARILUOMA[1,a], Matthias RAUTERBERG[2,b]

[1]Cognitive Science, University of Jyväskylä, Finland

[2]Industrial Design, Eindhoven University of Technology, The Netherlands

[a]pertti.saariluoma@jyu.fi; [b]g.w.m.rauterberg@tue.nl

***Abstract***

Many important lines of argumentation have been presented during the last decades claiming that machines cannot think like people. Yet, it has been possible to construct devices and information systems, which replace people in tasks which have previously been occupied by people as the tasks require intelligence. The long and versatile discourse over, what machine intelligence is, suggests that there is something unclear in the foundations of the discourse itself. Therefore, we critically studied the foundations of used theory languages. By looking critically some of the main arguments of machine thinking, one can find unifying factors. Most of them are based on the fact that computers cannot perform sense-making selections without human support and supervision. This calls attention to mathematics and computation itself as a representational constructing language and as a theory language in analysing human mentality. It is possible to notice that selections must be based on *relevance*, i.e., on why some elements of sets belong to one class and others do not. Since there is no mathematical justification to such selection, it is possible to say that relevance and related concepts are beyond the power of expression of mathematics and computation. Consequently, Turing erroneously assumed that mathematics and formal language is equivalent with natural languages. He missed the fact that mathematics cannot express relevance, and for this reason, mathematical representations cannot be complete models of the human mind.

***Keywords***

*Computation; Consciousness; Formal Language; Mind; Model; Turing machine*

## Preface

This paper is of a programmatic nature. We fully acknowledge the enormous achievements of modern science, engineering and design by calling on the most advanced machine models, as we always did in the past and will continue into the future. We will not primarily discuss the physical limitations (Markov 2014), but instead focus on the conceptual constraints, the intuitive assumptions of underlying theories, resp. their foundations (Saariluoma 1997). We try to understand and find an answer to the fundamental question: Is the human mind capable of understanding itself beyond computability? We question the mainstream assumption that everything is (or at least should be) 'computational' (Chatelin 2012, Chalmers 1996, Sun, Wilson, and Lynch 2016). We argue that the foundations of such an assumption are (still) not fully justifiable. Therefore, we imitate Kant's (1781/1922) famous 'Copernican Revolution' from a kind of Wittgensteinian (1921/1974) perspective and ask whether the properties of the *theory of language* itself used in discourse can explain why the problems have proven to be so hard. In other words, we ask *whether* formal theory languages (i.e., logic, mathematics and computation) are powerful enough to express problems of human thinking and represent thoughts. Since many of the foundational issues concentrate on one theoretical construct, Turing machines (TM), we must once again consider whether people 'think like machines'.

## Introduction

We are approaching the time of 'strong' artificial intelligence (AI) (Searle 1980). Machines can perform increasingly complicated tasks to free people from them (Sheridan 2002). Computers can beat people in chess, and in 20 years, cars and airplanes may drive or fly themselves (Brynjolfsson and McAfee 2014). However, there are many related, philosophically motivated issues (Hofstadter and Dennet 1981) that are conceptually unclear, such as whether people are machines in a sense, whether human intelligence is similar to computing, whether computers *think* like people and whether they can exert power over people (Horst 2003). This paper explores these issues by

investigating the intuitive assumptions and tacit limitations underlying the crucial concepts (Saariluoma 1997).

FIGURE 1. MACHINE METAPHOR TO DESCRIBE THE HUMAN MIND [ADAPTED FROM (LIGHTSPRING 2014)]

For a long time, already the most advanced machine model was used to capture and describe the most recent understanding of the human mind. After the mechanical watch was invented, the human mind was described as a complex mechanical machine (see FIGURE 1). Since the mid of last century, the computer took over as the dominant metaphor. Probably the idea of quantum effects based on entanglement will be used in the near future (Rauterberg 2008).

During the last century, Alan Turing created the hypothetical Turing Machine (TM). In 1936, he constructed to model how mathematicians think (Turing 1948/1996). The TM has a read-write head and an infinitely long squared tape. Each square can contain symbols '0' or '1'. The head can move one step to the left or right, and it can rewrite or leave the symbol as it is on the square when it is scanning. The head moves following pre-established transition rules that determine what the system does next. If there is no applicable rule, the system stops (Monk 1976).

Turing (1936) assumed that the TM models the thinking of a mathematician, and generalized the idea to all human thinking (Turing 1950).The symbols in machines could be organized into 'states of mind' and manipulated using a finite set of well-defined operations. Thus, Turing (1950) understood human thinking as a rule defined by the manipulation of symbols. The TM is familiar to most cognitive scientists. It is the basic notion behind computer science, the philosophy of computing, AI and modern cognitive science.

Turing (1936, 1948/1996) showed that the machine can do all of the (Turing computable) computations, and thus it can model the 'thinking of the mathematician', called the Hilbert program (Hilbert 1917). The thought process proceeds as follows. The machine is in an initial state, and then it scans a symbol and responds to it using a relevant transition rule. Consequently, the machine moves to another state. Again, it scans the symbol and applies to the next state. This process continues until the machine has either reached a satisfactory end state or has no more rules to apply. The transition functions and symbol patterns on the tape form a complete representation of the action. For example, symbols on the tape can represent a chess position, while transition functions are chess moves representing possible actions. The crucial property of a TM is, thus, that it seems to represent reality in some sense like human thinking. All of the directions seem to link to the idea that human information processing is in some sense similar to that of computers, but one cannot claim to have definitively determined whether machines think like people (Toni et al. 2007). Nevertheless, the discussion illustrates that there are different ways to develop the idea of an intelligent and thinking computer, i.e., a computational theory of the mind and beyond (Froese and Ziemke 2009, Horst 2003, Siegelmann 1995, Aarts et al. 2006, Nelson 1969).

In the field of the philosophy of the mind, important theoretical constructs include the computational theory of the mind (Putman 1964)and the representational theory of the mind (Fodor 1987, 1983). These ways of thinking perceive mentality as computation: i.e., they assume that the operation of the human mind can be expressed in terms of computing. Consequently, the TM models what the mind does. It is still open to debate whether the TM is a good model of the mind, and whether it really expresses all aspects of human mentality.

In any case, the TM is in many senses at the centre of the discussion about whether machines think and are intelligent in the same way as people. The TM is the core concept of modern computing and the basic model of computational thinking (Aho 2012, Wing 2006). It is also the first mathematically sound formal model to address human thinking (Turing 1950). This notion also embodies the concept of thinking as an algorithmic process, which uses a finite and definite set of operations or steps to lead from an initial state to a goal state. This notion is considered to be independent of the underlying computational power. The TM was central to developing modern cognitive psychology, which assumed that external human behaviour could be explained in terms of internal states (Neisser 1967/2014, Newell, Shaw, and Simon 1958, Simon and Newell 1956). It also inspired other cognitive research fields such as transformation grammar in linguistics (Chomsky 1957/2002, 1965/2014), cognitive neuroscience (Churchland and Sejnowski 1992) (Kosslyn and Andersen 1992), cognitive anthropology (D'Andrade 1995), psychology (Gigerenzer 2004), sociology (Luhmann 1984), art research (Nakatsu et al. 2015) and design thinking (Dorst 2011). In cognitive science, the TM has been central to discussions of the nature of human thinking (Newell and Simon 1972, 1976, Penrose 1999). Thus, the TM (and its many derivative concepts) has been the main conceptual tool in modern attempts to replace people with machines in performing certain tasks (Brynjolfsson and McAfee 2014).Turing was also able to develop a number of important solutions such as 'problem solving' as searching in a multi-dimensional state space (Newell and Simon 1972) and learning machines (Nilsson 1965), which showed that machines can carry out human-like intellectual tasks (Copeland and Proudfoot 1996). Later these findings led to important insights into machine learning, machine linguistics, logic computing and neural networks (Russell and Norvig 1995).

Nevertheless, it has been very hard to believe that machines *think* like people; the thesis of machine thinking has attracted several critiques from the very beginning (Dreyfus 1972, 1992) (Gunderson 1964) (Putnam 1964) (Searle 1990) (Wittgenstein 1953/1967). Many researchers did not and do not still accept the idea that machine information processing was an accurate model of human thinking. Yet we can ask: how are machines different from people?

At the same time, practical life has proceeded in its own ways. Much of modern human work is based on human intelligence, which is in turn based on the idea of the computational rationality of the human mind. Automation has changed industry in many ways as machines have replaced people in many routine tasks. Yet people still carry out apparently simple looking tasks because it has not been possible to build machines to do them. Thus, one of the most important developments in modern society is to construct technologies that can perform such tasks. Nevertheless, in order to design more intelligent machines, it is essential to clearly understand the similarities and differences between human and machine intelligence and thinking. The machine intelligence discussion has been

ongoing for many decades (Michalski, Carbonell, and Mitchell 1984, Copeland, Posy, and Shagrir 2013, Megill, Melvin, and Beal 2014); the issues must somehow be conceptually blurred (Brooks et al. 2012). The problem is how and why – and, of course, how the discourse could (and should) be resolved. To understand whether machines 'can think like people', the notion of TM must be subjected to critical conceptual analysis focusing on their foundations' (Saariluoma 1997, Saariluoma and Rauterberg 2015).

To do so, we adopt a new approach to the problem. We analyse the origins of the dispute by asking why it is still alive, and what makes it so difficult to determine (1) whether people think like machines and (2) whether human thinking can be used to model the mind. This kind of analysis, which aims to clarify scientific thinking, can be called foundational analysis (Saariluoma 1997). Here, we focus on one critical issue: what can be expressed by formal concepts; computational thinking is based on the assumption that formal models can express everything that human thinking can contain. We turn our attention from the ontology of computational models and human thinking to ask what kinds of language we use to analyse human thoughts when we use TMs and other formal constructions to analyse human thinking.

## Turing's Test

The debate over whether people think like computers was mainly intuitive and metaphorical. Therefore, it became important to find a critical model with which to test whether it makes sense to argue that people think like (and are intelligent in similar sense as) computers. To determine whether machines can think like people, Turing designed his famous test, which is likely the best-known thought experiment in AI (Besold 2013, Saygin, Cicekli, and Akman 2000, French 2012, Weinert 2014, You 2015); much of the man–machine dispute revolves around it.

### *Main Purpose of Turing's Test*

The goal of Turing's Test (TT) (Turing 1950) is to determine whether machines can think (or, as the question is often expressed, whether machines are intelligent). The TT is an imitation game (Warwick and Shah 2014). It is assumed that there is an opaque screen. On one side of the screen, an interrogator asks questions and assesses the nature of the answers. In the control group, on the other side of the screen, there are two people, A and B, who answer the questions. The interrogator must guess who has given the answer. In the experimental group, B is replaced by a machine (computer) and again, the interrogator must decide whether the human or the machine answered the questions. A teletypewriter was used to eliminate the problems caused by the quality of the voice communication (Turing 1950). Turing argued that if machines can imitate human thinking sufficiently perfect, then they are intelligent.

Thus, Turing's imitation game provides an explicit way to compare human and machine behaviours in intelligent tasks by abstracting from the physical embodiment (Pfeifer and Scheier 2001). Since discussion of the intelligence of machines underpins much of modern cognitive science, psychology and philosophy of the mind – and is also essential in developing AI robots and autonomous systems – it makes sense to consider the true value of the TT (Boden 1977). The decisive criterion in these experiments is the interrogator's capacity to determine whether a human or machine provided the answer. If the interrogator cannot do this above chance level, then the machine has passed the test and proven that machines can think, as they can perform human tasks in a way that it is indistinguishable from human performance by a competent observer.

It is obvious that the TT is a particularly clever idea, but in what way exactly? Generations of renowned researchers have considered all aspects of the test (Besold 2013, You 2015). The test was an innovation intended to justify Turing's computational theory of the mind (Turing 1936, 1950), but it was criticized even before its publication – and even before the presentation of the TM (Wittgenstein 1935/1958, Shanker 2002).[1]

Since the TT was published, it has been discussed by many notable researchers such as Newell and Simon (1959, 1972), Dreyfus (1972), Dennett (1998) and Searle (1980) who have ended to precisely opposite positions. Therefore, it is unsurprising that the test is still considered somewhat elusive (Saariluoma and Rauterberg 2015). On the one

[1] Wittgenstein's *The Blue and Brown Books* were finished between 1933 and 1935, before Turing's 1936 paper.

hand, it is easy to identify machines that can beat people in their particular fields, for example, chess machines and pocket calculators. On the other hand, no general man-like computer Leviathan exists (Dretske 2013) (Dreyfus 1992) (Searle 1990). Therefore, the conceptual analysis of the foundations of TM must be extended to TT. In this way, it would be possible to reach conceptual clarity around all computational thinking (Saariluoma 1997). In order to analyse the conceptual properties of TT, it is first worth considering some of the major critical arguments against the idea that machines can be engaged in human thinking, or that computational machines can implement all human information processes (Rauterberg 2008, Robinson 1998).

### *The Problem with Turing's Test*

Turing's original paper specified that he meant to apply the test to decide – either in particular cases or generally – whether machines can be as intelligent as people. For machines to be generally as intelligent as people, they must be able to take care of any imaginable action as efficiently as human beings. Thus, they would be considered as a *universal or all-purpose TM.* For example, Deep Blue, when it beats Kasparov in chess, it has proved to pass a specific TT, but of course its victory did not mean that TMs could be better than people at managing a multinational company. Thus, the difference between specific and universal TT is very clear. It will be possible to say that computers think like people only if an all-purpose TM can be developed.

Another important confusion in Turing's test was due to the era in which he worked. Psychology was then behaviourist, which meant that psychologists were interested in finding laws to describe how external stimuli led to external responses (Watson 1914). Presumably for this reason, Turing (1950) did not pay attention to internal machine processes that produce human-like external behaviours. Thus, Deep Blue had little in common with people. It searched for hundreds of thousands of positions in a second, while people normally study approximately 50 positions in ten minutes (De Groot 1965) (Saariluoma 1995). Thus, despite the similarities in external behaviours, the internal processes are very different.

Of course, the goal of most machines is to be *more effective than* people. It would not make sense for an excavator to be as effective as a man with a spade. This is why the TT is a very practical conceptual tool with which to assess whether technologies perform better than people, i.e., whether they take care of the things that they should take care of (Saariluoma and Rauterberg 2015). Well-performing machines do not need to have anything to do with human intelligence. They just take care of data processes that were previously only possible for people. Consequently, the TT only explores whether special purpose machines can do what people do, and in this it is a vital conceptual tool. It is important to note here that the TT is an empirical test: whether a chess machine is as good as a world champion is an empirical question.

## Critiques of Machine Thinking

The idea of the human as a computing machine has raised a number of critical counterarguments, which are important in analysing the scope and limitations of computational thinking (Leibniz 1714/1969). Turing (1950) addressed the main critiques of the time when presenting his thought experiment. First we discuss the Gödel critique (Lucas 1961), as it forms the foundation of most of the other critiques.

### *Gödel's Critique*

In the 1930s, distinguished scholars such as David Hilbert, Alan Turing, Kurt Gödel, Alonzo Church and others conducted fundamental work in mathematics that established the theoretical foundations of computability. Their work provides precise characterizations of effective and algorithmic computability, and can be seen as providing deep insights into the foundations of mathematics and calculability. Above all, it implicitly seeks the limits of mathematical concepts and mathematics as a representational language. In an unpublished manuscript, Turing (1948/1996, p. 256) wrote: 'By Gödel's famous theorem, or some similar argument, one can show that however the machine is constructed there are bound to be cases where the machine fails to give an answer, but a mathematician would be able to. On the other hand, the machine has certain advantages over the mathematician.' These advantages are still seen as the deterministic nature of such machines: with the same input, the calculations

produce exactly the same output. Mathematicians dream of a formal system that always provides the correct answer, is absolutely reliable and will not suffer from a 'mechanical breakdown' (Copeland, Posy, and Shagrir 2013, Megill, Melvin, and Beal 2014). Even Gödel (1961/1995, p. 377) was convinced that:

> "[M]athematics, by its nature as an *a priori science*, always has, in and of itself, an inclination toward the right, and, for this reason, has long withstood the spirit of the time [Zeitgeist] that has ruled since the Renaissance; i.e., the empiricist theory of mathematics, such as the one set forth by Mill, did not find much support. Indeed, mathematics has evolved into ever higher abstractions, away from matter and to ever greater clarity in its foundations (e.g., by [giving] an exact foundation of the infinitesimal calculus [and] the complex numbers)-thus, away from scepticism."

This conviction is based on the belief that such a super rationality is achievable and desirable. Furthermore, Gödel was also convinced that mathematicians, as human beings, are capable of this objective based on the *unlimited power* of the human mind.

Since Gödel started to challenge Turing's argument that 'machines can think' (Turing 1950), the whole debate has concentrated on the question of whether a TM is an adequate model of the human mind. The concept was first stated by Turing (1936), and can be summarized as follows: any formalizing process can be reproduced by a machine capable of performing an ordered series of operations on a finite number of symbols. Gödel (1972/1990, p. 306) wrote:

> "Turing in his 1936-1937 paper, (latter part page 250), gives an argument which is supposed to show that mental procedures cannot go beyond mechanical procedures. However, this argument is inconclusive. These problems are serious as Turing disregards completely the fact that mind, in its use, is not static, but constantly developing, i.e., that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding."

This quote demonstrates that Gödel considered the human mind to be bigger than any possible machines due to its constant but irreversible changes (Shotter 1974, Rauterberg 2013). His main argument is his understanding that the human mind is irreversibly changing and must be therefore open and very likely unlimited. Gödel's crucial finding was that there are obviously true theorems in any formal system that cannot be proven within the system they come from (Gödel 1931). This means that the language of proof cannot *express* all possible theorems. To do so, one would need to construct a new language.

Indeed, in constructing a special TM, people have to select the appropriate axioms and transformation rules to reach the goal. This thinking is beyond the abilities of any TM. Megill, Melvin and Beal (2014, p. 82) are still trying to argue against this position by assuming that 'the set of mathematical truths that humans know at any given moment' is finite, which Gödel never considered relevant. Gödel published little, but what he did publish still holds, particularly the two incompleteness theorems (Hodges 1998).

Does Gödel's idea make any sense? It is possible to construct a proof of any theorem, but not a system that can prove *all* theorems. It is thus possible to write a simulation model for any human thought, but not for *all* possible human thoughts. If we assume that all real numbers belong to set of human thoughts, it is possible to develop an algorithm that calculates this real number or human thought. It is possible to ascribe a value to pi or the biggest prime so far, but finite machines cannot calculate the accurate value of pi or the biggest of all primes. This means that a TM should decide what is a good enough value or contents for these numbers or human thoughts. Evidently, defining 'good enough' is beyond mathematical thinking. This conclusion means that we have at least two human (mathematical) thoughts that go beyond the capacity of any TM. Only people can solve these problems.

Of course, it is possible for humans to say that defining pi with 143 decimals is good enough, and that 145 decimals are not necessary. No machine can set this problem, as the idea of 'the right value of pi' is not expressible in any formal system. The language of the mathematics of formal logic is not powerful enough to express such problems. They can be discussed only in terms of natural language. Additionally, there are also many thoughts that are not mathematical at all.

***Informality of Behaviour***

The final conclusion of Gödel's critique makes some essential critiques of machine thinking understandable. A point that is close to Gödel's problem is Turing's 'informality of behaviour argument' (Turing 1950, p. 452). The core of this argument is that there are no rules that can mechanically explain all human behaviour and thinking. People may stop when they see a green light, but they also may not. They may change their behaviour. Thus, human behaviour cannot be fully determined by a fixed and finite set of rules in the way that machine behaviour can. The argument is linked to, though not the same as, the *Robots dilemma* (Pylyshyn 1987, Ford and Pylylshyn 1996), which holds that any change in an environment requires a change in the system of rules a robot uses. Another closely linked problem is the *frame problem*, which refers to the fact that the actions of robots change the environment and thus make it fluent (Hayes 1973). Some actions do not change any essential environmental properties, while others do. Therefore, the set of rules that controls the behaviour of the robots should be constantly updated.

The core of all these discussion is the *rules* that control the behaviour of machines and minds. One should know when a rule should be applied, and when it should not. It is essential to ask whether the existing set of rules is sufficient to always direct the behaviour (or thinking) of the machine in an efficient manner. Informality and rule system arguments have two practical consequences that have long been well known in both AI and cognitive simulation: exponential growth and conflict resolution. Exponential growth refers to the fact that the size of search systems tends to grow exponentially (Minsky 1961). Conflict simulation refers to the problem that in rule-based systems, more than one rule can often be applied in a decision situation. Rule systems thus have the problem of finding the best applicable rule – a *rule application problem* (Dreyfus 1972). So any rule-based system that solves tasks relevant for thinking seems to have problems applying rules and generating new rules when necessary. This means that, again, the problem of selection seems to be outside machine thinking. Machines cannot tell us what the sense-making rules are. Gödel proved that this is the case with any axiomatic formalism. Humans can intentionally (or even un-intentionally) break such rules (Dahling et al. 2012). Sometimes we are even encouraged to break rules to move forward. We break the rules, if we see it important or relevant. For people, rules of thought and action are no mechanical and fixed regulators, but regularities we follow flexibly (Buanes and Jentoft 2009).

***Lady Lovelace's Argument***

Wittgenstein (1935/1958) often emphasized that people not only *see* things; they 'see them as something'. This problem of 'seeing as' is deeply related to any consideration of whether machines can think (Wittgenstein 1935/1958). The first critical argument following from 'seeing as' was undoubtedly presented by Lady Lovelace (Turing 1950). Her argument was focused on Babbage's analytical engine, which was one of the first versions of a 'thinking' machine. She claimed that such engines could not initiate anything, especially anything new. Turing (1950) could not accept this argument, but did not present very convincing counterarguments. If her critique means that modern computers cannot randomly generate anything, it is not correct: computer programs can generate many things; the problem is whether the things they 'initiate' make any sense. Thus, we reformulate the requirement to mean 'anything that makes sense'.

Computers can generate a huge amount of results, but they cannot really determine what makes sense. An old British Museum algorithm suggested that ten apes with mechanical typewriters could eventually generate Shakespeare's collected works (Newell, Shaw, and Simon 1958). The problem is that they do not know where to stop, and in this sense they cannot even *start* the process. They should be able to re-create the works from among all other possible alternatives in order to be able to initiate and finalize the solutions. As is normal in creativity research, we interpret Lady Lovelace's argument to mean that she supposed the outcome of initiation would be something *essential and meaningful*.

To be able to initiate something, one must be able to define the goal and how to achieve it (i.e., the actions that lead from the initial state to the selected goal state). This is possible only if one is able to separate the required path of concepts from ones that are not required. One must be able to separate the goal-congruent and non-congruent aspects of the environment. The machine should thus be able to set a goal and the congruent activities

independently of its creator.

Setting the goal and generating goal-congruent activities can be called intentionality; actions generated in this way can be called intentions (Fishbein and Ajzen 1975, Bratman 1987). Initiating things, 'seeing as', selecting the goals and congruent actions, and acting following the generated representation are necessary goals to form a cluster of integrated activities, which seems to be very difficult for machines to achieve. For people, the actions of this cluster are a normal part of life. As for this cluster, counterarguments belongs the argument that computers, in contrast to people, lack 'free will'. Computers are determined by mechanical processes, and thus they cannot choose their own goals or make real choices. 'Free will' is only possible if one can set one's goals and intentions freely. One can thus select between one's deeds and initiate deeds at will. Initiating deeds and goals, and selecting between courses of action, is in some sense a sign of free will.

Of course, it is, to some degree, debatable whether people have free will or whether it is only an illusion (Libet, Freeman, and Sutherland 2000) caused by an inability to know the true mechanisms behind one's deed, as Leibniz (1714/1969) claimed. Nevertheless, it is evident that the human ability to make 'freely willed' choices is different from that of machines. Humans can set goals and meet them. They can (and do) initiate deeds. But computing machines must always eventually be activated and controlled by people. Humans must still define the sense-making goals and thus perform the selections. Thus, Lady Lovelace's argument is deeply connected to the notion of sense-making selection (Louis 1980, Schoenfeld 1992, Kurtz and Snowden 2003).

## Consciousness

It is not well known that Ludwig Wittgenstein (1953/1967, 1921/1974) was an opponent of machine thinking. However, he was perhaps the first to pay attention to the problems of using machines as models of human thought. He knew Turing well, and was familiar with his ideas (Monk 1990, Shanker 2002), and was thus well informed about the human dimensions of TMs. Wittgenstein's early work on logic preceded Turing's thinking, as he argued that logic defines the *apriori* limits of human thinking (Wittgenstein 1921/1974). Nevertheless, Wittgenstein (1935/1958, 1939/1975) put forward a number of critiques of Turing's idea that people are computers. For example, people feel pain, while machines do not (Wittgenstein 1937/1976). This apparently almost trivial remark touches on three aspects of the man–machine dispute. First, machines (either symbolic or mechanical) have no biological sensitivity to pain in the same sense that people have. Second, machines do not have emotions like people; and third, they cannot be conscious of pain. It is perhaps best to begin with the last aspect of Wittgenstein's. His conclusion was straightforward: machines cannot be thinking creatures (Wittgenstein 1953/1967, Nyíri 1989).

The pain argument makes sense only if one is conscious of pain. This is why we also have to consider the consciousness problem, of which Turing was very much aware. Of course, many other theorists have since thought about the problems of consciousness (Dreyfus 1992, French 1989, Koch and Tononi 2011). Turing (1950) simply bypasses the phenomenal aspects of experience in his argumentation and points out that TMs can behave as if they were conscious. For example, they can write sonnets. One important tacit assumption behind this argument over consciousness is that the phenomenon is seen as a single and un-analysable whole. Indeed, we do not have one single conceptual framework with which to investigate consciousness (Tononi and Koch 2015). Therefore, it may be easily misleading how the problem of consciousness should be dealt with in a theoretical discussion about the issues of thinking computers (Floridi 2005). Instead of directly discussing the hard problem, we can consider the preconditions that make consciousness possible – one of which is to generate sense-making contents for consciousness (Kurtz and Snowden 2003, Schmidt 1990).

In *Minds, Brains and Programs* and elsewhere, John Searle (1980, p. 417, 1990) argues against the thesis that 'computers, given the right programs, can be literally said to understand and have other cognitive states'. Searle clearly does not accept Turing's claim that a computer can think, because thinking (like feeling) presupposes conscious and intentional representations. People know what they think about; they do not merely mechanically process the contents of the information.

Consciousness can be understood by means of unconscious processes (Rauterberg 2008, 2010). Thus, on a biological level, it is possible to block the consciousness of pain by using painkillers. Thus, people have a pain-like state of mind, but no consciousness of it. They could be like zombies and thus equivalent to machines. Machines can process information, but they are not aware of the outcome. When a human chess player stares at a chessboard, she or he may be aware of many possible moves but equally ignorant of many others (Saariluoma 1995, Bilalić and McLeod 2014). One chess player can have a very clear idea about what is a good move, but another may not. Thus, consciousness also requires being able to conceptualize visual information. This dimension of consciousness can be called cognitive consciousness. It can be seen as the capacity to produce mental states that are normally conscious in the human mind.

This theme has been important to many phenomenological-oriented researchers. Dreyfus (1992), for example, argued that machines do not become conscious in same sense as people, but that they can behave externally as if they were conscious (i.e., like a conscious person). Thus, one can ask whether machines have the capacity to generate information states that are typical of human thinking (Buttazzo 2001). This is the point in Turing's argument regarding the consciousness dispute (Turing 1950). There is no doubt that there can be TMs that can compose music or generate poems. However, it is unclear whether they select the axioms as processes that lead to a poem. Can they decide by themselves what, out of all the possible combinations of words, is a good poem, unless people have selected it? This rather pessimistic view of TMs and machine thinking is a consequence of Gödel's argument.

Consciousness introduces an additional aspect to the discourse about how accurately TMs model the human mind. Evidently, people are not conscious of everything they see. For example, doctors and patients interpret X-ray images differently. Patients are mostly not conscious of the kinds of issues that are highly meaningful for doctors. Therefore, it makes sense to ask what someone can be conscious of. Assuming that TMs can be conscious of certain things, one can ask whether there are things they can be unconscious of.

This problem can be turned into the problem that machines cannot be conscious of issues that they cannot represent in their 'minds'. Thus, to answer the main question – whether there are things that machines cannot be conscious of – we have to show that there are things that are representable by people, but not by machines. Fortunately, the world is not in want of such issues.

Good examples are infinity and eternity. People can easily use these concepts, but they are not representable by machines. The problem of how many decimals of pi make sense provides an additional example. Machines can represent the 1567$^{th}$ decimal of pi, which may be very hard for people, but they cannot tell whether its relevance is different from the 1673$^{th}$ decimal. Machines can be told what is relevant, but they cannot determine it for themselves. Two prominent examples from the past can demonstrate this: (1) Deep Blue was the first machine, which defeated a human in chess (Campbell, Hoane, and Hsu 2002); this is a dedicated chess playing machine, which decides for the next move through the implemented algorithms based on the available computational power. (2) Alpha GO beats the most advanced human Go player; the success of this dedicated machine is based on the huge data set making it could act as us for its fully automated training process; the computational power and the high speed created a data set of played Go games that is far above all Go games ever played by all humans at all times (Wang et al. 2016). But does it mean that these two machines can understand the differences between chess and go? No, but people do!

One may remark that chess or go machines think like real human players. There is no way to separate their playing from human players based on the shown behaviour only. This means that these machines pass TT. However, TT is mistaken in two senses. Firstly, it does not consider the process of generating output and thus the way game playing machines 'think' is very different from how human chess players think. The machines generate huge internal search spaces while people perform very selectively. Thus, if there were a 'cognitive TT' (i.e., one which compares the information processing between man and machines), then such programs could not pass. Secondly, programs operate only in well-defined and narrow environments. They do not select the legitimate goals, operators or data elements, but these important pieces of information are given by programmers. Thus, human

thinking surpasses machine thinking in these respects. From this point of view, relevance and relevance-based selectivity is again a result of human thinking only.

## Representational Relevance Problem

The critiques of machine thinking are not unrelated, though they look at human and machine information processing from different points of view. Machines cannot be conscious: they can show or communicate emotion like expressions, but they do not know what those are. Emotions are central when people define the value of things to them. Therefore, it is understandable that computers cannot initiate things – or, to be precise, they cannot initiate the *right* things. Moreover, without human aid, computers cannot select what is essential. They can apply rules, but they cannot know why they apply these rules, as they do not have goals or intentions. Computers are not able to select their goals, or the rules they follow. Consequently, they are also unable to determine which parts of the environment belong to a meaningful focal area and which do not. All these pieces of knowledge must be programmed and thus defined for computers. In short, one can say that computers have great difficulty in deciding what is *relevant* or *irrelevant* without explicit human supervision. One can call this as the *relevance problem*. Whether the Non-Axiomatic Reasoning System introduced by Xu and Wang (2012) can resolve this issue is unclear.

One must ask what 'relevant' means in this context. Sperber and Wison (1986), who investigated the concept in linguistics, brought this theme to cognitive science. Here, we apply this problem to the TM context. As TMs process symbols and their combination following a set of rules, the notion must be defined in these concepts. Finally, it should also be based on the fundaments of formal systems, which means in set theory. Indeed, taking the notion of set as the basic concept in defining relevance makes sense, because set theoretical concepts underlie the notions of TM.

Sperber and Wilson (1986) suggested that relevance should be discussed in terms of *the principle of relevance*, which separates relevant and irrelevant issues. In the TM context, this would mean a principle of determining which elements in a set of symbols are relevant or irrelevant – for example, which values of pi are relevant. They discussed the need for a principle one could use to divide any set into relevant and irrelevant set members. Without such a principle, they cannot be separated; the relevance principle is not determinable in formal and mathematical concepts – rather, it is a mental concept.

Based on their evolutionary history, humans have this capability to understand *relevance*. Any species' phenotype is based on its genotype, which has evolved over many generations. Each species has a build-in basic set of needs, requirements and desires (NRD) (Salem, Nakatsu, and Rauterberg 2009). Those NRDs are structurally coupled with the living environment, called *autopoiesis* with the understanding of cognition as a biological process (Maturana 1975). Autopoiesis can only take place in the molecular domain of living systems, even when such living systems also exist as organisms in supra-molecular spaces (Maturana Romesin 2002). The NRDs are supposed to guarantee the survival of the offspring, and enable even the phenotype to develop higher needs on top of basic survival needs (Maslow 2013). From beginning of our life, this basic set of inborn NRDs determines, what is relevant and what not, to guide our actions and thinking. This structural coupling becomes reflected in our mental and emotional content structure.

Humans' understanding of *relevance* is grounded in their *biological* existence. The content of humans' mental representations is relevant, but people must always define for technical artefacts (e.g., computers), what is relevant or irrelevant. Even if computers can have internal states, and these internal representations may have their own contents, it is that such content is only relevant or irrelevant with respect to what those computers have to do. The relevance problem, when associated with representations, can be called a*re presentational relevance problem* (Dreyfus 2007). Representational relevance explains why computers do not have emotions, intentions, goals, selectivity, consciousness or a capacity to initiate actions. People must define these properties for machines in working AI programs. The question remains why *relevance* is such a problem for TMs (Dreyfus and Wrathall 2009).

***Power of Expression***

Ludwig Wittgenstein observed that human languages (or the concepts in them) do not form a single whole. Instead, they are formed from an unlimited number of sublanguages that he termed language games. The meanings of terms or symbols are defined in language games, and therefore have meaning only in the context of these games. If a concept does not belong to a language game, it has no real meaning. In *The Blue and Brown Books*, for example, he wrote, 'the trouble is rather that the sentence, "A machine thinks (perceives, wishes)" seems somehow nonsensical. It is as though we had asked "Has the number 3 a colour"' (Wittgenstein, (1935/1958).

Colour is a meaningless concept in mathematics, and therefore it is odd to say that the number 3 has a colour. Thus, the very question is meaningless in some way, because machines and thinking apparently belong to different language games, and therefore we should get some clarity about the very concepts of thinking and machine. Of course, the colour of a number may make sense, for example, in designing commercial banderols. Scientific theory languages are also language games, and therefore the meanings of terms must be defined in that context (Kurtz and Snowden 2003). However, the example illustrates that it is essential to analyse language, which is used to build mental representations and discuss them. One can call this issue as the problem of theory language, i.e., the problem of what kinds of concepts we should use to represent reality in scientific theories and intelligent machines (Xu and Wang 2012).

Scientific theory languages do not always cover what they should. For example, it is impossible to find a natural number that expresses such issues as the relationship between the side and the diagonal in a square or the value of pi. To represent such entities, it is essential to use real numbers. Similarly, in behaviourist psychology, it is not possible to consider the properties of human memory or mental images, as they are more than simply stimulus–response concepts. Finally, if we had restricted ourselves to Dalton's concept of the atom, modern nanophysics would also have been impossible. In science, progress is always about finding new concepts and exceeding the limitations of old ones (Bunge 1967, Kuhn 1962, Saariluoma 1997).

In essence, the *power of expression* is a notion that describes the scope and limits of a scientific theory language and its conceptual system (Bloom 1995). Thus, the power of expression defines and expresses what can be thought when a particular theory language is used. This notion can thus be applied to analyse the limits of formal computational languages. It also enables us to ask questions about the limits of computational concepts and therefore the real meaning of TT. TMs claim to be models of human thinking. As process models, they also have the capacity to perform tasks. Their performance naturally depends on how good the models of the constructed machines are(Simon and Newell 1956), which in turn is based on how well they can solve the representational relevance problem. The better they are at behaving like people, the more relevant their representations are, and the better they can perform the task. However, the level of their performance depends on the quality of their representations. Since TMs are constructed using mathematical theory language, it makes sense to discover the limits of mathematical language games in modelling thinking. One can also formulate these problems differently and ask how well mathematical theory language can represent human thinking (Brooks 1991). Does it have limits that decrease its power of expression in representing human thinking? The critical problem seems to be selection. All critical discourse seems to raise this point in one form or another (Bartos 2015).

In mathematical concepts and the meta-science of mathematics, *relevance* can be seen as the rule that determines which elements of any mathematical set (of elements or functions) are relevant vs. irrelevant. In terms of TMs, one should be able to say which combination of zeros and ones is relevant and which is not. Machines should also be able to determine which rules and functions are relevant and which are irrelevant. They should also be able to say, what is the right functional form needed in specific tasks.

In order to understand the reason for the problem of *relevance* and its connection to the sense-making selections, one must examine the notion of representation. TMs can behave like humans, as they can in some sense represent reality like people do (in the sense that they can generate human-like behaviours) (Simon and Newell 1956). Pocket calculators, for example, can do arithmetic and represent the number and arithmetical operators. Thus, the contents of their representations give machines the capacity to behave as people behave. Therefore, one can give the

relevance problem another form. The question is actually how relevant the representations are presenting the context of operating. Thus, one could also discuss the *representational relevance problem* (Xu and Wang 2012).

The power of expression suggests a straightforward question: can formal (mathematical or logical) theory languages express relevance? If they cannot, it is impossible to have a mathematical machine that could be relevant but in numerical contexts. Thus, no TM could be relevant in real-world contexts. A simple way to analyse the conceptual foundations of mathematical systems is to ask whether there is a mathematical way to decide which elements or functions of a mathematical set are relevant and which are not. It is obvious that mathematical theory languages cannot be used to decide the relevance or irrelevance of set members. It is always necessary to use non-formal theory languages. Consequently, we never know outside the narrow scope of mathematical contexts whether a mathematical representation is relevant.

The representational relevance problem is unsolvable in formal representations, as the power of expression for formal languages is too poor. One cannot use formal languages to discuss qualities and their nature (Saariluoma 1997). TMs and mathematical models of human thinking are constructed by abstracting semantic and thought content (Simon and Newell 1956). Consequently, one can no longer present what is relevant in a concrete context. There is no mathematical means to say that the formula 3+4 = 7 concerns carrots rather than tomatoes. To discuss tomatoes and carrots, one needs to have a language that can express qualities and make differentiations between relevant and irrelevant contents.

The formal relevance problem also has a role in formal logic. It is valid to infer that 'Napoleon was the Emperor of France' from the factually true sentences: 'the moon is not cheese', and 'if the moon is not cheese, Napoleon was the Emperor of France.' Whether the inference makes any sense is another issue. It is impossible to find any relevant connection between Napoleon's role as the Emperor of France and the fact that the moon is not cheese. The given facts have no relevant associations, and therefore the inference is formally valid but senseless.

Searle (1980) also noticed an important aspect of the formal relevance problem: syntax cannot generate semantics. This is why the origins of meaning giving must be sought in human conceptualization and judgment processes, which is why they are outside the framework of computational modelling. When the information contents of propositions are abstracted, it is impossible to produce sense-making, real-world semantics (Rauterberg 2006). It is impossible to define what is true and what is false, or what is right and what is wrong in the real world. This means that only interpretation in terms of real-world concepts, i.e., programmed semantics, makes it possible for AI systems to have any relevance. However, it is not possible to create semantics or to solve formal relevance problems on the basis of formal syntax only.

As Turing's idea of computation was originally based on his idea of a mathematical mind, and he extended his interest from mere mathematics to general computation, from early on the TM and its many derivatives became important in philosophy and the psychology of the human mind. Today, it is very common to think that human brains perform computations and that computational models may shed light on problems such as how people think and what intelligence is (Crane 2003, Fodor 1987, Newell and Simon 1972, Fodor 2000). Turing simply assumed that the manipulated symbols need not be numbers; they could just as well be Chinese characters or words in a natural language:

> "Thus, an Arabic numeral such as 17 or 999999999999999 is normally treated as single symbol. *Similarly*, any European language words are treated as single symbols (Chinese, however, attempts to have an infinity of enumerable symbols). The difference in view between the single and compound symbols is that the compound symbols cannot be seen at one glance…"(Turing 1936) (emphasis added).

These symbols can represent 'states of mind', which can be operated by a machine. Consequently, Turing (1950) argued that machines could be intelligent, and that they could in some sense think. Thus, like many others after him, Turing assumed that the TM was, in some senses, a model of the human mind (Newell and Simon 1959). However, this paragraph entails a serious error, as it loosely equals mathematical symbols and words in natural languages.

However, Turing did not specify how the real-world symbols and their meanings were associated with the TM (Hodges 1998). The associations between the number combinations on the tape, the symbols and the references to the symbols are given, but are not sufficiently processed (Steels 2008, Taddeo and Floridi 2005, Müller 2015). Thus, the most important action in human thinking (i.e., determining what is relevant vs. irrelevant) is omitted from the computational modelling. As shown above, formal languages lack sufficient power of expression to analyse how computational models can be combined with reality. So far, our considerations have focussed on the limited power of mathematical and other formal concepts in expressing human thinking. However, it is easy to present thousands of working AI systems, and thus it is logical to ask how working AI can be explained. Obviously, their real-word semantics work fine. Thus, it is possible to develop fine AI, even though formal semantics has only a limited power of expression. Consequently, it is essential to ask how this is possible (Müller 2015).

TMs cannot be models of the human mind or human thinking, as they are built on formal concepts. TMs are formal mathematical systems, which is why they cannot express information contents. The terms in abstract languages are void of meaning. Therefore, they cannot be used to discuss such real-world issues as *relevance*, *truth* or *value*. Nevertheless, many AI systems can solve these problems and they work well. To solve this expression, we have to think about the problems of the power of expression from a new point of view.

In formal systems, the notions of meaning, semantics and truth are based on model theoretical semantics, which means that the truth of any proposition is determined by comparing the proposition with a model set, which has members with no reference to the external world. Thus, the truth of a proposition is defined by comparing its terms with a given mathematical set. The meaning and sense of expressions in model theoretical semantics depends on their truth, i.e., propositions only have meaning if they have truth values.

TMs or any intellectual machines need not be based on abstract theoretical models only. They can also be built on concepts that can be interpreted in the real world. A computational device following human movements in the environment and reacting to different states can be built only by interpreting the meanings of sensory data. Hand movement, for example, can mean the use of a tennis racket in a computer game. The movement is assigned to the logic of the program, and thus the meanings in a machine have a real-world interpretation.

***Symbol Grounding***

Computational concepts can be used to build representations. They can be words, propositions, pictures or signs. The main thing is that they have their references in the real world. They represent, or stand for, their references. Thus, the references form the basis for the meanings of the symbols. Yet, this description does not solve the problem of meaning. One must also ask how symbols get their meanings – i.e., how they are connected with their references. This problem is called as *symbol grounding* (Harnad 1990); it may appear as a unified problem, but it is not. There are several different perspectives to investigate in order to determine how symbol grounding is possible. Many researchers are interested in the neural basis of symbol grounding. However, this perspective normally sets the question on a rather high level so that the research does not focus on the precise meaning (or mental contents) of the symbols.

In principle, in TMs symbols get their meaning and information contents from their references. However, symbol grounding is not a single-layer issue. Today, computer vision can be used to collect masses of optical tracking points about sports, for example. However, the third point mass makes sense only if it is possible to find sense-making patterns inside. Of course, this is precisely what TMs are about. They are tapes of ones and zeros. The combination of these can be used to present symbols so that one set of points can be seen as one symbol and another set as another symbol. The crucial point is why some sets of ones and zeros mean one thing, let's say the Chinese symbol for happiness, and others can be read as tripling in basketball.

'What Turing disregards completely is the fact that *mind, in its use, is not static, but constantly developing*, i.e., that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding' (Gödel 1972/1990, p. 306). In this quote, it becomes clear that the process of meaning giving is crucial in applying computational thinking, and it is the biggest difference between

human and machine information processing. For example, a machine does not know what 'checkmate' is unless a programmer has defined and programed it. Is there a kind of machine learning possible, which is comparable to humans' capabilities?

## Turing Demystified: Mental Contents, Relevance and Artificial Intelligence

Despite their intuitive clarity, the concepts of TM and TT make tacit conceptual commitments that must be critically considered (Gurevich 2012). Thinking that mathematical symbols form similar symbol systems as natural languages was a serious error. It simply led to the incorrect idea that human and machine information processing are the same. Clearly, there are similarities (Gigerenzer 2004, Toni et al. 2007), but they should not hide the important differences. The first sign of a problem is Gödel's argument that no universal TM can be constructed. The second problem is that there is no meaningful mathematical method to choose the right TM from among different possible TMs using mathematical arguments only. TMs can neither generate meaningful goals of action by themselves nor know what meaningful representational contents are. Or as Floridi (2010, p. 404) summarizes, "that—given the current state and understanding of computer science—the best artefacts that artificial intelligence (AI) will be capable of engineering will be, at most, zombies: artificial agents capable of imitating an increasing number of human behaviours."

All the problems discussed above converge into one central notion –*relevance*. Mathematics cannot tell us which parts of the elements of a set belong to the right subset and which do not. All the subsets of any set are equally important in a mathematical sense, but not in the real world. Thus, formal languages cannot be used to discuss such central properties of human thinking as the ability to know what is true, what is right and what is beautiful. They are all notions that make sense only if relevance can be expressed.

When we speak about truth, it is essential to know what is relevant. A proposition can be true only if it has a real reference. However, knowing what belongs to the reference presupposes the ability to decide what kinds of things belongs to the referred. Of course, deciding what is a car is possible only if one knows the relevant properties of the object. To know what is right or beautiful presupposes the ability to decide what properties and objects are relevant.

Our analysis illustrates that Turing did not accurately formulate his idea about computers as thinking machines. He also did not differentiate between special and all-purpose TMs. However, his most serious error was assuming that mathematical and natural languages use qualitatively similar signs and numbers. Mathematical TMs cannot think like people do, and cannot even model human thinking, because formal theory languages cannot express qualitative differences.

Computational machines are special-purpose TMs. This means they can perform real-world actions such as text processing, image analysis and temperature control. Modern AI is continuously developing more interesting computational systems. Autonomous systems, which, broadly speaking, can, to some degree, redefine their goals during operation, are a good example of the capacity of emerging technologies. However, a merely intuitive interpretation of computational systems does not make sense. It is time to go beyond the limits of computational concepts and admit that we need a new way of thinking that incorporates computational representations of information contents.

The TT is vital, as it can be used to evaluate the performance of computational systems. If they can perform as well as (or even better than) people, they are intelligent in a computational sense. That is good enough, as special-purpose computational machines can essentially improve the quality of human life. Another problem is that they can model human thinking, but they cannot think like people. They cannot express relevance – this is defined by human programmers.

If it is possible to develop context-specific fixed semantics, it is possible to construct a special-purpose machine to replace human work. However, it is impossible to construct an all-purpose machine or universal TM that can think and speak like a human being as long as we have to use fixed-theory languages. One might argue that it is possible

to construct a universal TM that calculates the functions of all TMs and takes import of all of them. Unfortunately, this solution does not work. This kind of machine would have an unlimited number of TMs. Any time a new functionality occurs in the world, it would be necessary to construct a new machine. This construction game would be endless, and the notion of endless is itself a problem.

Let's assume that we could have a 'TM of all TMs'. Could it any longer be a TM? The set of all functions of all TMs and the set of all inputs of all TMs would have to bypass Russell's paradox. They would be sets of all sets, but could they be sub-machines of themselves any longer? Thus, it seems that mathematics is not a sufficiently powerful theory language to represent what people think. We can simulate human thinking – and even replace it in any context – but we cannot argue that formal languages can represent *everything* that people can think.

Finally, one may ask, what the origin of relevance problem is. The answer is straightforward: formal languages are reached by abstraction. This means they are devoid of real-world contents, which form the basis of our natural languages and their semantics. Reverse thought operations (i.e., concretization and going from abstract to concrete concepts) make it possible to discuss relevance, because relevance is a content-level issue. The problem makes sense only with concrete issues, but this means that any TM must eventually be constructed using non-formal intuitions.

There is no problem generating random sets of symbols and random constructions using a TM. The problem is known, which of the sets of symbols makes sense. An engineer can have a set of all parts of a jumbo jet, and by a trial-and-error search can precisely find the right combination of parts. Yet he has to be competent enough to say which of the endless possible combinations of parts is correct – which is impossible on the grounds of formal knowledge only.

We have ended with a strange conclusion: the problem of human and machine intelligence and thinking is an illusion created by the poor power of expression that is typical of formal languages. It makes sense to ask the problem only in the context of discourse based on formal theory languages. People can ask what is relevant, and they can set themselves relevant goals. To be able to discuss such issues, one needs to use another type of theory language. Undoubtedly, phenomenological discourses provide a good example of languages that can be used to investigate issues outside the sphere of formal discourse (Dreyfus 1972, Heidegger 1927/1962, Husserl 1900-1901/2008, Merleau-Ponty 1945/1996). We have to consider humans as growing, evolving and developing systems based on irreversible processes which can – if at all – be understood in a holistic manner (Capra and Luisi 2014).

## ACKNOWLEDGEMENTS

This research is partially funded by the Industrial Design Department of Eindhoven University of Technology and Design United (Netherlands). We are very thankful to Rene Ahn, LoeFeijs, and an anonymous reviewer for their valuable feedback on earlier versions of this article.

## REFERENCES

[1] Aarts, Emile, Herman Ter Horst, Jan Korst, and Wim Verhaegh. 2006. "Computational intelligence." In *True Visions*, edited by Emile Aarts and Jose L Encarnação, 245-273. Berlin - Heidelberg: Springer.

[2] Aho, Alfred V. 2012. "Computation and computational thinking." *The Computer Journal* 55 (7):832-835.

[3] Bartos, Vıt. 2015. "Biological and artificial machines." In *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, edited by Jan Romportl, Eva Zackova and Jozef Kelemen, 201-210. Cham - Heidelberg - New York - Dordrecht - London: Springer.

[4] Besold, Tarek Richard. 2013. "Turing revisited: A cognitively-inspired decomposition." In *Philosophy and Theory of Artificial Intelligence - SAPERE 5*, edited by Vincent C. Müller, 121-132. Berlin: Springer.

[5] Bilalić, Merim, and Peter McLeod. 2014. "Why good thoughts block better ones." *Scientific American* 310 (3):74-79.

[6] Bloom, Lois. 1995. *The Transition from Infancy to Language: Acquiring the Power of Expression*. Cambridge: Cambridge

University Press.

[7] Boden, Margaret Ann. 1977. *Artificial Intelligence and Natural Man*. Vol. 5. Hassocks: Harvester Press.

[8] Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.

[9] Brooks, Rodney A, Demis Hassabis, Dennis Bray, and Amnon Shashua. 2012. "Is the brain a good model for machine intelligence?" *Nature* 482 (23 February):462-463.

[10] Brooks, Rodney A. 1991. "Intelligence without representation." *Artificial Intelligence* 47 (1–3):139-159. doi: http://dx.doi.org/10.1016/0004-3702(91)90053-M.

[11] Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York - London: W. W. Norton.

[12] Buanes, Arild, and Svein Jentoft. 2009. "Building bridges: Institutional perspectives on interdisciplinarity." *Futures* 41 (7):446-454. doi: http://dx.doi.org/10.1016/j.futures.2009.01.010.

[13] Bunge, Mario. 1967. *Scientific Research*. Vol. I-II. Berlin: Springer

[14] Buttazzo, Giorgio. 2001. "Artificial consciousness: Utopia or real possibility?" *Computer* 34 (7):24-30.

[15] Campbell, Murray, A. Joseph Hoane, and Feng-hsiung Hsu. 2002. "Deep Blue." *Artificial Intelligence* 134 (1):57-83. doi: http://dx.doi.org/10.1016/S0004-3702(01)00129-1.

[16] Capra, Fritjof, and Pier Luigi Luisi. 2014. *The Systems View of Life: A Unifying Vision*. Cambridge: Cambridge University Press.

[17] Chalmers, Davide J. 1996. "A computational foundation for the study of cognition." University of California Accessed May 31. http://cogprints.org/319/1/computation.html.

[18] Chatelin, Françoise. 2012. *Qualitative Computing - A Computational Journey into Nonlinearity*. Singapore: World Scientific Pub.

[19] Chomsky, Noam. 1957/2002. *Syntactic Structures*. 2, revised ed. Berlin - New York: Mouton de Gruyter.

[20] Chomsky, Noam. 1965/2014. *Aspects of the Theory of Syntax*. Cambridge: MIT press.

[21] Churchland, Patricia Smith, and Terrence J. Sejnowski. 1992. *The Computational Brain*. Cambridge: MIT press.

[22] Copeland, B. Jack, Carl J. Posy, and Oron Shagrir, eds. 2013. *Computability: Turing, Gödel, Church, and Beyond*. Cambridge - London: MIT Press.

[23] Copeland, B. Jack, and Diane Proudfoot. 1996. "On Alan Turing's anticipation of connectionism." *Synthese* 108 (3):361-377.

[24] Crane, Tim. 2003. *The Mechanical Mind: A philosophical introduction to minds, machines and mental representation*. 2nd ed. New York: Routledge.

[25] D'Andrade, Roy G. 1995. *The Development of Cognitive Anthropology*. Cambridge: Cambridge University Press.

[26] Dahling, Jason J., Samantha L. Chau, David M. Mayer, and Jane B. Gregory. 2012. "Breaking rules for the right reasons? An investigation of pro-social rule breaking." *Journal of Organizational Behavior* 33 (1):21-42. doi: 10.1002/job.730.

[27] De Groot, Adriaan D. 1965. *Thought and Choice in Chess*. The Hague - Paris - New York: Mounton.

[28] Dennett, Daniel Clement. 1998. *Brainchildren: Essays on Designing Minds*. Cambridge: MIT Press.

[29] Dorst, Kees. 2011. "The core of 'design thinking' and its application." *Design Studies* 32 (6):521-532. doi: http://dx.doi.org/10.1016/j.destud.2011.07.006.

[30] Dretske, Fred. 2013. "Machines and the mental." *Proceedings and Addresses of the American Philosophical Association* 59 (1):23-33.

[31] Dreyfus, Hubert L. 1972. *What Computers Can't Do*. New York - Evanston - San Francisco - London: Harper & Row.

[32] Dreyfus, Hubert L. 1992. *What Computers Still Can't Do - A Critique of Artificial Reason*. Cambridge - London: MIT Press.

[33] Dreyfus, Hubert L. 2007. "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian." *Philosophical Psychology* 20 (2):247-268.

[34] Dreyfus, Hubert L., and Mark A. Wrathall. 2009. *A Companion to Phenomenology and Existentialism*. New York: John

Wiley&Sons.

[35] Fishbein, Martin, and Icek Ajzen. 1975. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading: Addison-Wesley.

[36] Floridi, Luciano. 2005. "Consciousness, agents and the knowledge game." *Minds and Machines* 15 (3-4):415-444. doi: 10.1007/s11023-005-9005-z.

[37] Floridi, Luciano. 2010. "The philosophy of information: Ten years later." *Metaphilosophy* 41 (3):402-419. doi: 10.1111/j.1467-9973.2010.01647.x.

[38] Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge: MIT Press.

[39] Fodor, Jerry A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge: MIT Press.

[40] Fodor, Jerry A. 2000. *The Mind does't Work that Way*. Cambridge: MIT Press.

[41] Ford, Kenneth M., and Z. Pylylshyn, eds. 1996. *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*. Norwood: Ablex.

[42] French, Christopher C. 1989. "The case against mental duality." *Current Psychology* 8 (3):200-218.

[43] French, Robert M. 2012. "Moving beyond the Turing test." *Communications of the ACM* 55 (12):74-77.

[44] Froese, Tom, and Tom Ziemke. 2009. "Enactive artificial intelligence: Investigating the systemic organization of life and mind." *Artificial Intelligence* 173 (3):466-500.

[45] Gigerenzer, Gerd. 2004. "Fast and frugal heuristics: The tools of bounded rationality." In *Blackwell Handbook of Judgment and Decision Making*, edited by D. Koehler and N. Harvey, 62-88. Oxford: Blackwell.

[46] Gödel, Kurt. 1931. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." *Monatshefte für Mathematik und Physik* 38 (1):173-198.

[47] Gödel, Kurt. 1961/1995. "The modern development of the foundations of mathematics in the light of philosophy." In *Collected Works (CW)*, edited by S. Feferman, JW. Dawson, W. Goldfarb, C. Parsons and RM. Solovay, 374-387. New York - Oxford: Oxford University Press.

[48] Gödel, Kurt. 1972/1990. "Some remarks on the undecidability results." In *Collected Works (CW)*, edited by S. Feferman, JW. Dawson, SC. Kleene, GH. Moore, RM. Solovay and J. van Heijenoort, 305-306. New York - Oxford: Oxford University Press.

[49] Gunderson, Keith. 1964. "The imitation game." *Mind* 73 (290):234-245.

[50] Gurevich, Yuri. 2012. "Foundational analyses of computation." In *How the World Computes*, edited by S. Barry Cooper, Anuj Dawar and Benedikt Löwe, 264-275. Berlin - Heidelberg: Springer.

[51] Harnad, Stevan. 1990. "The symbol grounding problem." *Physica D: Nonlinear Phenomena* 42 (1):335-346.

[52] Hayes, Patrick J. 1973. "The frame problem and related problems in artificial intelligence." In *Artificial and Human Thinking*, edited by A. Elithorn and D. Jones, 45-49. San Francisco: Jossey-Bass.

[53] Heidegger, Martin. 1927/1962. *Being and Time*. London: HarperOne.

[54] Hilbert, David. 1917. "Axiomatisches Denken [Axiomatic Thinking]." *Mathematische Annalen* 78 (1):405-415.

[55] Hodges, Wilfrid Augustine. 1998. "Turing's philosophical error?" In *Concepts for Neural Networks*, edited by L. J. Landau and J. G. Taylor, 147-169. London: Springer

[56] Hofstadter, Douglas R., and Daniel C. Dennet, eds. 1981. *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic Books.

[57] Horst, Steven. 2003. "The computational theory of mind." Stanford University Accessed 18 Dec. http://plato.stanford.edu/archives/spr2013/entries/computational-mind/.

[58] Husserl, Edmund. 1900-1901/2008. *Logical Investigations*. Edited by Dermot Moran, *International Library of Philosophy*. London, NewYork: Routledge.

[59] Kant, Immanuel. 1781/1922. *Critique of Pure Reason*. Edited by Friedrich Max Müller. 2nd revised ed. New York: Macmillan.

[60] Koch, Christof, and Giulio Tononi. 2011. "A test for consciousness." *Scientific American* 304 (6):44-47.

[61] Kosslyn, Stephen Michael, and Richard A. Andersen, eds. 1992. *Frontiers in Cognitive Neuroscience*. Cambridge: MIT Press.

[62] Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

[63] Kurtz, Cynthia F., and David J. Snowden. 2003. "The new dynamics of strategy: Sense-making in a complex and complicated world." *IBM Systems Journal* 42 (3):462-483.

[64] Leibniz, Gottfried Wilhelm. 1714/1969. "The principles of nature and of grace based on reason." In *Gottfried Wilhelm Leibniz: Philosophical Papers and Letters* edited by L.E. Loemker, 636-642. Dordrecht: Reidel.

[65] Libet, Benjamin, Anthony Freeman, and Keith Sutherland, eds. 2000. *The Volitional Brain: Towards a Neuroscience of Free Will*. Exeter: Imprint Academic.

[66] Lightspring. 2014. "Human brain with cogs and gears." Shutterstock, Inc. Accessed 15 June. http://www.shutterstock.com/pic-81976456/stock-photo-human-intelligence-with-grunge-texture-made-of-cogs-and-gears-representing-strategy-and.html.

[67] Louis, Meryl Reis. 1980. "Surprise and sense making: What newcomers experience in entering unfamiliar organizational settings." *Administrative Science Quarterly* 25 (2):226-251.

[68] Lucas, J. R. 1961. "Minds, machines and Gödel." *Philosophy* 36: 112-127.

[69] Luhmann, Niklas. 1984. *Soziale Systeme [Social Systems]*. Frankfurt am Main: Suhrkamp.

[70] Markov, Igor L. 2014. "Limits on fundamental limits to computation." *Nature* 512 (7513):147-154. doi: 10.1038/nature13570.

[71] Maslow, Abraham H. 2013. *Toward a Psychology of Being*. reprint from 1962 ed. New York: Start Publishing LLC.

[72] Maturana, Humberto R. 1975. "The organization of the living: A theory of the living organization." *International Journal of Man-Machine Studies* 7 (3):313-332. doi: http://dx.doi.org/10.1016/S0020-7373(75)80015-0.

[73] Maturana Romesin, Humberto. 2002. "Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition." *Cybernetics & Human Knowing* 9 (3-4):5-34.

[74] Megill, Jason L, Tim Melvin, and Alex Beal. 2014. "On some properties of humanly known and humanly knowable mathematics." *Axiomathes* 24 (1):81-88.

[75] Merleau-Ponty, Maurice. 1945/1996. *Phenomenology of Perception*. Translated by Colin Smith. Delhi: Motilal Banarsidass Publishers.

[76] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. 1984. *Machine Learning: An Artificial Intelligence Approach*. Berlin - Heidelberg: Springer.

[77] Minsky, Marvin. 1961. "Steps toward artificial intelligence." *Proceedings of the IRE* 49 (1):8-30.

[78] Monk, J. Donald. 1976. "Turing machines." In *Mathematical Logic*, edited by J. Donald Monk, 14-25. New York - Heidelberg - Berlin Springer.

[79] Monk, Ray. 1990. *Ludwig Wittgenstein. The Duty of Genius*. London: Jonathan Cape.

[80] Müller, Vincent C. 2015. "Which symbol grounding problem should we try to solve?" *Journal of Experimental and Theoretical Artificial Intelligence* 27 (1):73-78. doi: 10.1080/0952813X.2014.940143.

[81] Nakatsu, Ryohei, Naoko Tosa, Matthias Rauterberg, and Wang Xuan. 2015. "Entertainment, culture, and media art." In *Handbook of Digital Games and Entertainment Technologies*, edited by R. Nakatsu, M. Rauterberg and P. Ciancarini, 1-51. Singapore: Springer.

[82] Neisser, Ulric. 1967/2014. *Cognitive Psychology: Classic Edition*. New York - London: Psychology Press.

[83] Nelson, R. J. 1969. "Behaviorism is false." *The Journal of Philosophy* 66 (14):417-452. doi: 10.2307/2024129.

[84] Newell, Allen, John Calman Shaw, and Herbert A. Simon. 1958. "Elements of a theory of human problem solving." *Psychological Review* 65 (3):151-166.

[85] Newell, Allen, and Herbert A. Simon. 1976. "Computer science as empirical inquiry: Symbols and search." *Communications of the ACM* 19 (3):113-126.

[86] Newell, Allen, and Herbert Alexander Simon. 1959. *The Simulation of Human Thought*. Santa Monica: Rand Corporation.

[87] Newell, Allen, and Herbert Alexander Simon. 1972. *Human Problem Solving*. Vol. 104. Englewood Cliffs: Prentice-Hall.

[88] Nilsson, Nils J. 1965. *Learning Machines*. New York: McGrawHill.

[89] Nyíri, Kristóf J. C. 1989. "Wittgenstein and the problem of machine consciousness." *Grazer Philosophische Studien* 33/34:375-394.

[90] Penrose, Roger. 1999. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.

[91] Pfeifer, Rolf, and Christian Scheier. 2001. *Understanding Intelligence*. Cambridge: MIT press.

[92] Putman, Hilary. 1964. "Robots: Machines or artificially created life?" *The Journal of Philosophy* 61 (21):668-691.

[93] Putnam, Hilary. 1964. "Robots: Machines or artificially created life?" *The Journal of Philosophy* 61 (21):668-691.

[94] Pylyshyn, Zenon W, ed. 1987. *The Robots Dilemma: The Frame Problem in Artificial Intelligence*, *Theoretical Issues in Cognitive Science*. Westport, CT: Greenwood Publishing Group Inc. .

[95] Rauterberg, Matthias. 2006. "HCI as an engineering discipline: to be or not to be!?" *African Journal of Information and Communication Technology* 2 (4):163-184.

[96] Rauterberg, Matthias. 2008. "Hypercomputation, uncounsciousness and entertainment technology." In *Fun and Games*, edited by P. Markopoulos, 11-20. Berlin: Springer

[97] Rauterberg, Matthias. 2010. "Emotions: The voice of the unconscious." In *Entertainment Computing - ICEC 2010*, edited by H.S. Yang, R. Malaka, J. Hoshino and J.H. Han, 205-215. Heidelberg: Springer.

[98] Rauterberg, Matthias. 2013. "How is culture and cultural development possible?" In *Proceedings 4th International Conference on Culture and Computing*, edited by Kozaburo Hachimura, Toru Ishida and Naoko Tosa, 177-178. Tokyo: IEEE Computer Society's Conference Publishing Services.

[99] Robinson, Guy. 1998. *Philosophy and Mystification: A Reflection on Nonsense and Clarity*. London - New York: Routledge.

[100] Russell, Stuart J., and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.

[101] Saariluoma, Pertti. 1995. *Chess Players' Thinking: A Cognitive Psychological Approach*. London - New York: Routledge.

[102] Saariluoma, Pertti. 1997. *Foundational Analysis. Presuppositions in Experimental Psychology*. London: Routledge.

[103] Saariluoma, Pertti, and Matthias Rauterberg. 2015. "Turing test does not work in theory but in practice." In *Proceedings of the 17th International Conference on Artificial Intelligence - ICAI* edited by H.R. Arabnia, 433-437. Herndon: Mercury.

[104] Salem, Ben, Ryohei Nakatsu, and Matthias Rauterberg. 2009. "Kansei experience: Aesthetic, emotions and inner balance." *International Journal on Cognitive Intelligence and Natural Intelligence* 3 (2):18-36.

[105] Saygin, Ayse Pinar, Ilyas Cicekli, and Varol Akman. 2000. "Turing test: 50 years later." *Minds and Machines* 10 (4):463-518. doi: 10.1023/a:1011288000451.

[106] Schmidt, Richard W. 1990. "The role of consciousness in second language learning." *Applied Linguistics* 11 (2):129-158.

[107] Schoenfeld, Alan H. 1992. "Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics." In *Handbook of Research on Mathematics Teaching and Learning*, edited by D. Grouws, 334-370. New York: Macmillan.

[108] Searle, John R. 1980. "Minds, brains, and programs." *Behavioral and Brain Sciences* 3 (3):417-457.

[109] Searle, John R. 1990. "Is the brain's mind a computer program." *Scientific American* 262 (1):26-31.

[110] Shanker, Stuart. 2002. *Wittgenstein's Remarks on the Foundations of AI*. London - New York: Routledge.

[111] Sheridan, Thomas B. 2002. *Humans and Automation: System Design and Research Issues*. New York: John Wiley & Sons.

[112] Shotter, John. 1974. "What is it to be human?" In *Reconstructing Social Psychology*, edited by N. Armistead, 53-71. Harmondsworth: Penguin Books.

[113] Siegelmann, Hava T. 1995. "Computation beyond the Turing limit." *Science* 268 (5210):545-548.

[114] Simon, Herbert A., and Allen Newell. 1956. "The uses and limitations of models." In *The State of the Social Sciences*, edited

by L. D. White, 89-104. Chicago: University of Illinois Press.

[115] Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Vol. 142. Cambridge: Harvard University Press.

[116] Steels, Luc. 2008. "The symbol grounding problem has been solved, so what's next?" In *Symbols and Embodiment: Debates on Meaning and Cognition*, edited by Manuel de Vega, Arthur M. Glenberg and Arthur C. Graesser, 223-244. Oxford: Oxford University Press.

[117] Sun, Ron, Nick Wilson, and Michael Lynch. 2016. "Emotion: A unified mechanistic interpretation from a cognitive architecture." *Cognitive Computation* 8 (1):1-14.

[118] Taddeo, Mariarosaria, and Luciano Floridi. 2005. "Solving the symbol grounding problem: A critical review of fifteen years of research." *Journal of Experimental and Theoretical Artificial Intelligence* 17 (4):419-445. doi: 10.1080/09528130500284053.

[119] Toni, Roberto, Giulia Spaletta, CD Casa, Simone Ravera, and Giorgio Sandri. 2007. "Computation and brain processes, with special reference to neuroendocrine systems." *Acta Biomed* 78 (1):67-83.

[120] Tononi, Giulio, and Christof Koch. 2015. "Consciousness: here, there and everywhere?" *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370 (1668, no. 20140167):1-18.

[121] Turing, Alan M. 1936. "On computable numbers, with an application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (2):230-265.

[122] Turing, Alan M. 1948/1996. "Intelligent machinery - a heretical theory." *Philosophia Mathematica* 4 (3):256-260.

[123] Turing, Alan M. 1950. "Computer machinery and intelligence." *Mind* 59 (433-460).

[124] Wang, Fei-Yue, Jun Jason Zhang, Xinhu Zheng, Xiao Wang, Yong Yuan, Xiaoxiao Dai, Jie Zhang, and Liuqing Yang. 2016. "Where does AlphaGo go: from Church-Turing thesis to AlphaGo thesis and beyond." *IEEE/CAA Journal of Automatica Sinica* 3 (2):113-120. doi: 10.1109/JAS.2016.7471613.

[125] Warwick, Kevin, and Hemal Shah. 2014. "Good machine performance in Turing's imitation game." *IEEE Transactions on Computational Intelligence and AI in Games* 6 (3):289-299.

[126] Watson, John Broadus. 1914. *Behavior: An Introduction to Comparative Psychology*. New York: Henry Holt & Co.

[127] Weinert, Alexander. 2014. "The Turing Test." Accessed 14 December. http://alexanderweinert.net/papers/2014turingtest.pdf.

[128] Wing, Jeannette M. 2006. "Computational thinking." *Communications of the ACM* 49 (3):33-35.

[129] Wittgenstein, Ludwig. 1921/1974. *Tractatus logico-philosophicus*. London: Routledge.

[130] Wittgenstein, Ludwig. 1935/1958. *The Blue and Brown Books - Preliminary Studies for the 'Philosophical Investigation'*. Edited by Peter Docherty. Oxford: Blackwell.

[131] Wittgenstein, Ludwig. 1937/1976. "Cause and effect: Intuitive awareness." *Philosophia* 6 (3-4):409-425. doi: 10.1007/BF02379281.

[132] Wittgenstein, Ludwig. 1939/1975. "Lectures I-XXXI." In *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*, edited by R.G. Bosanquet and C. Diamond, 11-294. Chicago - London: University of Chicago Press.

[133] Wittgenstein, Ludwig. 1953/1967. *Philosophische Untersuchungen. Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Blackwell.

[134] Xu, Yingjin, and Pei Wang. 2012. "The frame problem, the relevance problem, and a package solution to both." *Synthese* 187 (1):43-72. doi: 10.1007/s11229-012-0117-8.

[135] You, Jia. 2015. "Beyond the Turing Test." *Science* 347 (6218):116-116.