Non-computing Machinery and Intelligence

Rod Smith
associative-ai.com
31 March 2019

Abstract

The research field of Artificial Intelligence (AI) seeks to program a digital computer with the algorithms of mind. The challenge has been to adequately understand the mind. Yet AI has two research topics: 1. the mind; 2. the machine. It adopts the computer science concept of the machine. John Searle's Chinese Room Argument (CRA) attack on AI assumes this view then strongly argues that it precludes the possibility of semantic mental content, a necessary content of mind. Objections to the CRA accept the view then variously argue that nevertheless the CRA fails (systems reply, Chinese gym, robot reply, ...). I argue that rather than being scientific fact, AI's understanding of the machine is theory, and the theory is error-ridden. While this has little effect on computer science, it is devastating for AI. Searle is right, no computation will ever think. AI needs a non-computing machine. Happily, it already has one. If only AI had the right theory of its machine, it would understand why.

INTRODUCTION

In the summer of 1956, a disparate group of researchers gathered at Dartmouth College, Hanover, New Hampshire, for an informal 2-month workshop in what would become the most exciting and challenging technical project since the wartime Manhattan Project centered at Los Alamos, New Mexico.

Unlike the Manhattan project, the newly christened undertaking would be carried out in relative peacetime. Yet like the Manhattan Project, it would be mainly military funded in the hope of reducing the political and financial cost of foreign military action.

That summer workshop took place over 60 years ago. It was the inauguration in the United States of America of the research project of Artificial Intelligence, or AI. The flyer announcing the workshop, distributed in 1955 by organizer John McCarthy, read:

DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories
August 31, 1955

> We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.[1]

[1] Available online at www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html in January 2019.

The assumptions and perspectives of the flyer reveal much about the foundations and future of the project. The assumption of precise description and simulation references the Church-Turing thesis, Turing machine, the computational theory of mind, and the computational theory of the digital computer, and it reiterates Turing's 1951 statement, "*...machines [digital computers] can be constructed which will simulate the behaviour of the human mind very closely*"[2]. The assumption of science ("*...a carefully selected group of scientists...*") removes the need to question concepts.

AI as science became a strong theme. Two decades later, Allen Newell and Herbert Simon's classic AI paper, "Computer Science as Empirical Inquiry: Symbols and Search", published in leading US Computer Science journal *Communications of the Association for Computing Machinery*, is just one of many examples.

Since the 1970s, AI has been taught in Computer Science departments of Western universities. Leading university AI text since the mid-1990s, Stuart J. Russell and Peter Norvig's, *Artificial Intelligence: A Modern Approach*, explains:

> "*This book is primarily intended for use in an undergraduate course or course sequence. It can also be used in a graduate-level course (perhaps with the addition of some of the primary sources suggested in the bibliographical notes). Because of its comprehensive coverage and the large number of detailed algorithms, it is useful as a primary reference volume for AI graduate students and professionals ... The only prerequisite is familiarity with basic concepts of computer science...*"

These two assumptions, computation and science, are discussed below, and it's concluded that both the computational theory of the machine and the avoidance of conceptual analysis were two specially consequential mistakes. They were key errors, they were made right at the start, and they set a flawed conceptual framework for the ongoing project.

The Manhattan Project entailed general scientific consensus about core atomic theory and the nature of sub-atomic matter. However in AI it soon became apparent that there was little agreement about the precise nature of the mind to the level of detail and expressed in the concepts needed for realization in the digital computer. For example, while AI's dominant theory of mind, computationalism, said the mind is an executing computation, it failed to indicate which computation.

During the over half century since 1956, it has slowly become apparent to the AI research community that the project's original goal – human-like general intelligence in an electronic digital computer – is a far more difficult prospect than originally thought, in part because the fundamentals of general intelligence are not clearly understood.

There were early indications of this in that theorists and practitioners including project founder Alan Turing were unable to explain sensory perception, not only a necessary aspect of human-like general intelligence, but the coalface, the causal interface between the external and internal realms. For instance, Turing's 1950 paper regarded by many as the mission statement of AI, "Computing Machinery and Intelligence", failed to even mention sensory perception. Failure to understand the principles of sensory perception, that is the semantic principles, the epistemological principles, the algorithmic principles, the causal principles, meant progress in that area could be only serendipitous.

---

[2] A. M. Turing (1951), "Intelligent Machinery: A Heretical Theory", in B. Jack Copeland (Ed.) (2004), *The Essential Turing*, OUP, page 472.

The great potential of early AI research soon attracted attention, both of detractors and supporters. Perhaps the now most well-known attack on the idea of a thinking computer was philosopher John Searle's 1980 Chinese Room Argument. This concludes that the fundamental nature of the digital computer precludes the possibility of it understanding the world or anything. The digital computer is the machine AI seeks to make intelligent. For various reasons including processing speed, quantity of uniquely addressable memory locations, and direct access memory ("random access memory"), no other type of device is available or envisaged.

Searle's attack on AI differs from others such as the frame problem in that it analyses the concept of the digital computer and concludes from this that the fundamental nature of the machine precludes the possibility of it having a human-like or any mind.

Thus Searle's attack is more devastating than others such as the frame problem (McCarthy and Hayes, 1969[3]) and related problem of combinatorial explosion (Lighthill, 1972[4]). For while these two present serious difficulties, it is possible to imagine improvement in computer processing speed and parallelism that might be largely or completely remedial. Whereas if the fundamental nature and essential character of the device is such that it could never understand its surroundings including language, improvements in speed, parallelism or anything are pointless.

In some circles, the CRA is considered the biggest challenge to AI's goal of programming a computer with human-like general intelligence. The field of cognitive science by and large embraces the computational theory of mind, as does AI. Co-author of the 1969 paper introducing the frame problem, Patrick Hayes, as reported by Stevan Harnad in 2001, proposed that the field of cognitive science should be redefined as "*the ongoing research program of showing Searle's Chinese Room Argument* (CRA) *to be false*"[5].

AI technology raises questions of ethics. AI ethics has attracted recent interest, and the possibility of the destruction of humanity by the intelligent machine is reportedly a concern at high levels. Yet there is great potential for good since most or even all human poverty and global pollution could be eliminated. History provides an indication in that the development of chemistry from the foundation of empirical alchemy offered the prospect of great human evil, and various British, German, American and other governmental embodiments of tribal evil realized this, mainly in the form of gases targeting human respiration and in bombing. But overall, there has been great benefit from chemistry, though now chemistry is polluting the planet by industrial and vehicular emission of greenhouse gasses, and with environmental plastic. Yet the intelligent machine may decide, or be imbued with the inclination to decide, to itself destroy humanity. One danger is that military AI goes berserk.

The AI research field, the practical target of Searle's Chinese Room Argument, largely ignores the argument. However, I argue that the CRA is extremely important. This is because the argument is unsound, and seeing why it is unsound indicates a better way to understand the electronic digital computer and how it might acquire human-like knowledge.

The CRA is unsound because Searle's conception of the digital computer (which he adopts from AI and which AI adopts from computer science and which computer science adopts in

---

[3] John McCarthy and Patrick J. Hayes (1969), "Some Philosophical Problems from the Standpoint of Artificial Intelligence", *Machine Intelligence 4*, pages 463-502.

[4] James Lighthill (1972), "Artificial Intelligence: A General Survey", prepared for the British Science Research Council and presented in July 1972.

[5] Quoted by Harnad in Stevan Harnad (2001), "Minds, Machines, and Searle 2: What's Right and Wrong about the Chinese Room Argument", in Mark Bishop and John Preston (Eds) (2002), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, Clarendon Press, pages 294-307.

part from Turing) contains significant error. The CRA concludes that no computer could think, but what the CRA means by "computer" is something quite different from the actual electronic device that AI seeks to make intelligent.

In other words, the CRA is a sort of *ignoratio ellenchi* – it misses the point. It argues against a position that was never put in the first place. The position put in the first place is that the machine called a digital computer could think. The CRA concludes that the machine described by Searle's conception of the device, a computational conception, could never think. This is very true. Such a device so described could never think. But Searle's conception fails to accurately depict the real machine. Specifically, the CRA premise, computers necessarily compute, is false.

This raises the question, Is respective non-computing by a "computer" (so called) conducive to a human-like general intelligence? If it is, AI will not only have gotten a most egregious monkey off its back (the CRA), but it will also have acquired a possibly better way to view the abilities of its machine.

OVERVIEW OF MAIN ARGUMENT

Of the attacks on the idea of a thinking computer, Searle's celebrated Chinese Room Argument is perhaps the most well known. In it, he uses key concepts of AI to attack AI, in my view extremely successfully – in terms of those concepts. His argument concludes that because of the fundamental nature of the computer, it is not possible that a computer could think.

Criticisms of AI typically focus directly or indirectly on its theories of mind or characteristics of computation. AI's concept of its machine is rarely questioned. Both AI and its critics regard AI's understanding of the digital computer as scientific fact, hence not theory, hence not open to debate. However, I argue that this understanding *is* theory, and what's more, it's bad theory.

AI's theory, and hence Searle's theory, of the electronic digital computer contains initially very difficult-to-see errors. While these don't affect comprehension of the machine as a tool for human use, and in fact derive from its human use, they are devastating when it comes to trying to understand the intrinsic device and how it might think.

The CRA concludes that the machine's fundamental nature prevents it ever understanding its world. But if Searle's conception of the computer is wrong, if his concept is not relevantly true to the actual device, then it does not follow that computers will never think.

ECONOMICS AND AI

Clearly, the advent of genuine AI will have vast economic repercussions, and AI research has a responsibility to consider these. One benefit of true AI is presumably that capitalism will be consigned to the dustbin of history by the intelligent machine. It's hoped that structural adjustments will produce a new and ethical economic reality. With no labor value component to market pricing, the cost of production will plummet along with the quantum of income needed for an adequate life, and the societal benefit of efficiently and quickly allocating resources in response to changing demand in the context of limited resources, could soon become relatively unimportant. And if the American corporation seeks to profit from AI-related intellectual property, China, Russia and others could simply use it without license, to their advantage.

Almost everything is dissolved in sea water. With zero or near-zero labor cost presumably many of those things could be viably retrieved. Though diamonds are not dissolved in sea water, and the natural jewel is opium to the evolved tribalism of status and plumage. The labor cost of mining or collection being at near-zero wouldn't be a factor in the pricing of diamonds, but who controls the kimberlite pipe presumably will be decisive. The ad-men will still manipulate demand, and the owners of the natural inputs to production will rule supreme. And who would need a trade union? Hence the eternal conflict between capital and labor will not be eternal anymore.

If Marx is right and economics determines politics, and if WWI and WWII were adjustments to the economics of the Industrial Revolution including the rise of the industrial bourgeoisie and reaction to its cruelty (the reaction being in part trade unionism), and to the inroads of both the *nouveau riche* and unionism into into the traditional power base of the agrarian elites, then with real AI, there's going to be big trouble. We all know Power happily starts wars to retain influence, wealth and prestige.

SEARLE'S ANALYSIS OF AI CONCEPTS

Because Searle uses relevant AI concepts and carefully discusses these in his Chinese Room Argument, I use as a starting point his description and analysis of the concepts, particularly those of the symbol, computation and the digital computer.

In 1992 Searle said:

> "*...we are in a peculiar situation where there is little theoretical agreement among the practitioners on such absolutely fundamental questions as, What exactly is a digital computer? What exactly is a symbol? What exactly is an algorithm? What exactly is a computational process? Under what physical conditions exactly are two systems implementing the same program?*"[6]

Given this, many views about what is a symbol, what is computation etc., are going to be to some extent wrong, unhelpful, misleading. But maybe a small number will be right in relevant respects. In his Chinese Room Argument, Searle seeks to be clear about what he means by key terms.

PROVENANCE OF THE CONCEPTS

There's quite a provenance to AI's concepts. Searle particularly adopts those of the computer and computation. (He needs to use the same concepts as AI in order to mount a legitimate attack.) AI in turn borrows concepts from the field of computer science. One could go a step further: computer science accepts core ideas of Alan Turing, in particular that of his 1936 Turing machine[7] and elements of his descriptions of the electronic digital computer made between 1946 and 1951[8].

---

[6] John R. Searle (1992), *The Rediscovery of the Mind*, MIT Press, page 205.

[7] A. M. Turing (1936), "On Computable Numbers, with an Application to the Entscheidungsproblem", available online.

[8] A. M. Turing (1945), "Proposed Electronic Calculator"; (1945) "Notes on Memory", (1946-7), The Turing-Wilkinson Lecture series", all in B. Jack Copeland (Ed.) (2005), *Alan Turing's Automatic Computing Engine*", OUP; (1946), "Proposal for Development in the Mathematics Division of an Automatic Computing Engine (ACE)" in B. E. Carpenter and R. W. Doran (1986), *A. M. Turing's ACE Report of 1946 and other Papers,* MIT Press; (1947), "Lecture on the Automatic Computing Engine"; (1948), "Intelligent Machinery"; (1950), "Computing Machinery and Intelligence"; (1951), "Intelligent Machinery: A Heretical Theory"; (1951), "Can Digital Computers Think?", all in B. Jack Copeland (Ed.) (2004), *The Essential Turing*, OUP; (1951) "Programmers' Handbook For Manchester Electronic Computer Mark II", University

As for the electronic design of the digital computer, Turing was intimately involved in early British efforts in the mid- to late-1940s and early 1950s while working at the National Physical Laboratory, London, the University of Manchester Computing Machine Laboratory, and elsewhere. In America, author of the classic 1945 computer science report, "*First Draft of a Report on the EDVAC*"[9], John von Neumann, was well aware during the early 1940s of Turing's 1936 paper, and von Neumann "repeatedly emphasized its fundamental importance"[10].

CONCEPT OF THE COMPUTER

The Turing machine as described by Turing in 1936 comprises three parts, a tape divided into squares, a read-write-erase head, and a motor that moves the head one square at a time over the tape (or alternatively the tape one square at a time under the head). There are five "simple" operations: Print - create a symbol then deposit it in the empty square beneath the head; Scan – identify the shape of the symbol if any in the square currently under the head; Erase – destroy the symbol if any in the square under the head; Left – move the head one square to the left; and Right – move the head one square to the right. There is also Halt – turn off the machine. And there is the part of the machine that isn't mentioned by Turing but which sequences the simple operations and can be thought of as wiring.

Searle in describing the electronic digital computer links the concept of the computer to that of the Turing machine:

> "*Alan Turing gave half a century ago ... the definition of the computer*"[11].
>
> "*A computer is by definition a device that manipulates formal symbols. These are usually described as 0s and 1s ... The inventor of the modern conception of computation, Alan Turing, put this point by saying that a computing machine can be thought of as a device that contains a head that scans a tape. On the tape are printed 0s and 1s*"[12].

Turing, in his 1950 paper and under the section headed "Digital Computers", establishes this link:

> "*Usually fairly lengthy operations can be done [by an electronic digital computer] such as "Multiply 3540675445 by 7076345687" but in some machines only very simple ones such as "Write down 0" are possible*"[13].
>
> And in 1947,
>
> "*...digital computing machines ... are in fact practical versions of the universal [Turing] machine*"[14].

---

of Manchester, errata sheet dated 13 March 1951.

9 John von Neumann, *First Draft of a Report on the EDVAC*, Moore School of Electrical Engineering, University of Pennsylvania, 1945, available online.

10 B. Jack Copeland (2004), "Compuptable Numbers: A Guide", in B. Jack Copeland (Ed.) (2004), *The Essential Turing: The Ideas That Gave Birth to the Computer Age*, OUP, page 23.

11 John R. Searle (1997), *The Mystery of Consciousness*, Granta Books, London, page 59.

12 John R. Searle (1997), *The Mystery of Consciousness*, Granta Books, London, pages 9-10.

13 A. M. Turing (1950), "Computing Machinery and Intelligence", *Mind*, Vol. LIX, No. 236, October 1950, available online, page 437.

14 A. M. Turing (1947), "Lecture on the Automatic Computing Engine", transcript of the lecture delivered on 20 February 1947 to the London Mathematical Society, in B. Jack Copeland (Ed.) (2004), *The Essential Turing: The Ideas That Gave Birth to the Computer Age,* OUP, page 383.

Turing's 1936 description (in the form of a table) of the essential Turing machine had it manipulating 0s and 1s[15] described as follows:

m-config. name	symbol	further operations	next m-config.
b	None	P0	b
	0	R, R, P1	b
	1	R, R, P0	b

"Manipulating" here means creating a symbol, destroying a symbol, or moving a symbol (moving the square containing the symbol) relative to the read-write-erase head. The Scan operation is "described" in the second column and identifies whether the symbol beneath the head is 0 or 1. The machine's internal actions, these are described in the three rows of the table, differ depending on the shape of the scanned symbol. R means move the head (or tape) one square to the right, P1 means print a 1 (create a 1 then deposit it) in the square beneath the head, likewise P0. The symbol "b" is the name of what would now be called the routine, and this routine calls itself as described in the far-right column. The tape is initially empty.

As indicated earlier, AI's concept of the electronic digital computer, and therefore also Searle's, can be traced back to the Turing machine. A relevant point is that in introducing his idea of the Turing machine, Turing says,

> *"A machine can be constructed to compute the sequence 010101.... The machine is ... capable of printing '0' and '1' ... The behaviour of the machine is described in the following table..."*[16]).

This table is the above m-config (machine configuration) table.

However, I'd like to point out that this table is not entirely a description. When Turing says the machine is capable of printing a 0 and a 1, he is not describing what the Turing machine can print. The shapes 0 and 1 are not names of what it can print. They don't refer to or denote what the machine can print. They don't describe. Rather, they are reproductions of the very printed shapes. "Zero" and "one" are the names.

So the shapes 0 and 1 in the table don't refer. They mimic. They are another instance of the same thing. The machine manipulates objects of shape 0 and 1, and the so-called "description" of the machine simply includes new instances the same shapes: 0 and 1. The "P" in the table is indeed descriptive and means the print function of the head. But the "0" and "1" in the table are not descriptive. They replicate.

This will be a key point for later discussion. It is one of the various fundamental differences between a Turing machine and an electronic digital computer. It's a fundamental semantic difference. In a description of a digital computer, "0" and "1" are actually names. They don't replicate. They describe.

---

[15] The machine, once switched on, prints: 0 1 0 1 0 1 0 1 .... The space (empty square of the tape) between each 0 and 1 is not explained by Turing, but introduces the idea of interstitial initially-empty squares that may later be used for workings, which is relevant to his description of the universal machine.

[16] A. M. Turing (1936), "On Computable Numbers, with an Application to the Entscheidungsproblem", available online, page 233.

Anyone in AI who takes the picture of the Turing machine to be an accurate depiction of the electronic digital computer, who thinks that the things electronic digital computers process are semantically viable, have meanings as do the things Turing machines and the Chinese room processes, well, that would be a serious mistake. Neither the Turing machine nor the Chinese room embody the semantics of the electronic digital computer.

CAN COMPUTERS NON-COMPUTE?

Both AI and Searle say electronic digital computers when operating according to their electronic design necessarily compute. Searle about his Chinese room thought experiment: "*I simply behave like a computer; I perform computational operations...*"[17].

I specially want to ask whether this machine, for historical reasons called a "computer", could operate other than by performing computations. If it seems it could, then relevant questions are: (a) what sorts of non-computing could a computer perform? and (b) could any such sort be relevant to, or even necessary to, intelligence?

In other words, sure, Searle might be right – no computation will ever think – but what if computers could do things other than compute? These other things by themselves, or in combination with computing, might be sufficient for the digital machine mind.

DEFINING "COMPUTING"

Key to this approach is adequately defining "computing", a notoriously variously defined concept ranging from pancomputationalism (the universe is a computational system) to Turing's 1936 explanation: "*Computing is normally done by [a human] writing certain symbols on paper...*"[18]. The object of a definition is to either include or exclude a clearly articulated type of process. If we want to say that certain results come from using the process as defined, then the type of process is included. If we want to conclude that certain results could never come from using the process as defined, then the sort is excluded. Either way, research options are narrowed and more effective use of resources indicated.

REASON FOR AI'S MISTAKES

If AI has inadvertently made significant conceptual mistakes, we naturally want to understand why this has occurred, and ask, Is there an overarching reason? AI has adopted concepts of computer science. I argue that AI has made significant conceptual errors, and the overarching reason is that it has adopted the computer science concept of the digital computer. The field of computer science deals with the human use of the machine, and the concepts it uses to understand the device reflect this perspective. For AI, however, the idea of human-defined machine behavior is inappropriate.

AGI (the research field of artificial general intelligence) seeks to develop the intrinsic machine – the device in and of itself – and to develop it into a self-aware and independent cognitive entity that perceives its world and learns from it. The intelligent machine is not an automaton forever mindlessly executing the instructions of the human programmer. The idea of computing is actually a compound idea that derives from a human activity taught in schools. Different concepts, more fundamental ones, more accurate ones, are needed to understand how the intrinsic machine will perceive and understand its world.

---

[17] John R. Searle (1980), "Minds, Brains, and Programs", in *Behavioral and Brain Science*, 3 (3), 1980, pages 417-457, reprinted in John Haugeland (Ed.) (1981), *Mind Design*, MIT Press, page 285.

[18] A. M. Turing (1936), "On Computable Numbers, with an Application to the Entscheidungsproblem", available online, page 249.

NEED TO UNDERSTAND THE CONCEPT OF COMPUTING

But the idea of computing, or a relevant idea of computing, needs to be clearly delimited in order that other processes can be identified as being other processes.

COMPUTING AS SYMBOL MANIPULATION

Searle says computing is a process of manipulating symbols on the basis of their shapes. When talking about what is manipulated, he uses the term "formal symbol", and "*...all that 'formal' means here is that I can identify the symbols entirely by their shapes*"[19]. Further, computers necessarily compute, "*I simply behave like a computer; I perform computational operations...*"[20], and "*The computer operates by manipulating symbols*"[21], also "*A computer is by definition a device that manipulates formal symbols*"[22].

Of course one could define "symbol" to mean whatever computers process. However, this seems suspiciously like a form of begging the question. A way to avoid this mistake is to go back in time to before the computer and ask What is computation and what is a symbol? In 1936, Alan Turing explained: "*Computing is normally done by [a human] writing certain symbols on paper...*"[23].

In this human case, the input and output are objects, symbols, external to the human, such as numerals and names of variables. For example, converting currency using an exchange rate can be done manually with pen and paper by writing down numerals (currency amounts) and names of variables (currency designations) and following simple rules about symbols.

A key point is that the symbols, the interpretable shapes, the numerals, the variable names, exist outside of the human brain. No one seriously suggests that perceiving the external shape A means that necessarily something shaped A exists inside the skull, or that things inside the skull are manipulated on the basis of their shapes.

MACHINE CONCEPTION OF COMPUTING – INTERNAL MANIPULATION

The concept of machine computation now strikingly diverges from that of the human case. In the machine conception, the machine, in order to do anything, internally manipulates symbols. Searle: "*A computer is by definition a device that manipulates formal symbols*"[24] and "*...all that 'formal' means here is that I can identify the symbols entirely by their shapes*"[25]. Searle means internal manipulation. The machine is designed to internally detect and react to the shapes of the internal symbols, and that reaction is the base causation that allows the machine to operate.

It's not that the human conception expressly precludes this possibility of internal manipulation of symbols, interpretable shapes. It simply says nothing about it. The possibility is not part of the concept. But the idea of machine computation necessarily

---

[19] John R. Searle (1980), "Minds, Brains, and Programs", in *Behavioral and Brain Science*, 3 (3), 1980, pages 417-457, reprinted in John Haugeland (Ed.) (1981), *Mind Design*, MIT Press, page 284.

[20] John R. Searle (1980), "Minds, Brains, and Programs", in *Behavioral and Brain Science*, 3 (3), 1980, pages 417-457, reprinted in John Haugeland (Ed.) (1981), *Mind Design*, MIT Press, page 285.

[21] John R. Searle (2004), *Mind: A Brief Introduction*, OUP, page 63.

[22] John R. Searle (1997), *The Mystery of Consciousness*, Granta Books, page 9.

[23] A. M. Turing (1936), "On Computable Numbers, with an Application to the Entscheidungsproblem", available online, page 249.

[24] John R. Searle (1997), *The Mystery of Consciousness*, Granta Books, page 9.

[25] John R. Searle (1980), "Minds, Brains, and Programs", in *Behavioral and Brain Science*, 3 (3), 1980, pages 417-457, reprinted in John Haugeland (Ed.) (1981), *Mind Design*, MIT Press, page 284.

requires manipulating inner symbols. This is the point: in the human case, what is manipulated is external symbols. In the machine case, what is manipulated is internal symbols.

Searle about internal symbols:

> "*...a digital computer is a syntactical machine. It manipulates symbols and does nothing else.*"[26]; "*...the computer operates by manipulating symbols...*"[27]; "*A computer is by definition a device that manipulates formal symbols*"[28].

WHAT IS A SYMBOL?

Computing as traditionally understood by AI is fundamentally symbol manipulation according to rules about symbol shape. Hence it is quite relevant to consider what is a symbol. In the human case of computing, a symbol is a token, for example an instance of the substance of ink soaked into paper, whose shape has a meaning – it's a tokenized interpretable shape. And it's the shape that has the meaning, not the substance that bears the shape. Expressed in slightly more detail, the value of the property of shape has the meaning. The values are the different shapes. This is the linguistic-computational conception of the symbol. Values of other properties can also have meanings to humans. The colors of lenses in traffic signals, is one example.

In the machine case, the symbols are processed inside the device. The symbols inside a Turing machine and inside the Chinese room, an analogue of the digital computer, are tokenized interpretable shapes, 0s and 1s or Chinese ideograms. They are manipulated on the basis of their shapes (different shapes are treated in different ways). This is the linguistic-computational conception of the symbol.

However, this is where the computational concept of the computer radically differs from the actual machine. In the electronic digital case, the internal things called "symbols":

(a) are not processed on the basis of their shapes (the machine by virtue of its design and manufacture is incapable of reacting to the shapes, and hence the shapes are irrelevant to the causation of the machine); and

(b) have no meanings in the sense that linguistic-computational symbol shapes have meanings.

MEANING OF A SYMBOL

This lack of machine reaction to shape, again, is a departure from the case of human computing where the shapes of tokens are reacted to and have interpretations. The electronic digital computer reacts to voltage level, switch state, and magnetic orientation, not to shape. It reacts to values of these other properties rather than to values of the property of shape.

Yet this machine reaction is still reaction to values of a property of the things the machine internally "manipulates". So this is the key semantic question: In the human case it is the shape that has the meaning. If the properties are different in the machine case (and they are), do the values of these other properties also have meanings?

---

[26] John R, Searle (2014), "What Your Computer Can't Know", in *The New York Review of Books*, October 2014.

[27] John R. Searle (2004), *Mind: A Brief Introduction*, OUP, page 63.

[28] John R. Searle (1997), *The Mystery of Consciousness*, Granta Books, page 9.

MEANING IN THE MACHINE CASE

No – not in the sense that symbol shapes have meanings. Symbol shapes are perceived by humans, and we say the human assigns a meaning to the shape (which is how people learn languages). But humans can't perceive the values of the things computers internally process. So humans can't assign meanings to them.

One way to view this is as follows. We say humans assign meanings to perceived shapes. This establishes a 2-term relation. One term is a shape and the other, a meaning (ignoring the issue of universals). But what really happens is that the human connects (associates) an inner meaning with the inner *representation* of a shape gained through perception. (So now there are two particulars, and the issue of a universal being a term of a relation does not apply.) The relation is really between two internal things, an inner meaning and an inner representation of a shape, not between an inner meaning and an external shape. And the relational connection is physically realized in neural fiber. Both the meaning and the representation of the shape are inside the human brain.

Searle's picture is that the only semantics relevant to the machine is the meanings of the shapes of the symbols that he says the machine internally manipulates. He says the machine's internal processing could never access these meanings, which are inside the heads of human observers.

We can now see that this picture is flawed. Computers don't and can't internally react to the shapes of the data units they receive. So what is inside the human skull and linked to inner representations of shapes is totally irrelevant to understanding the semantics, or lack of semantics, of the computer.

The machine is designed to react to values of properties of the things it processes: clocked voltage level, switch state, magnetic orientation. Searle rightly says that values of properties have no intrinsic meaning. But he's wrong to say that the things computers process have an extrinsic meaning. Certainly, shapes can have extrinsic meaning, and they do in the case of text understood by humans. But humans can't perceive clocked voltage levels, etc. Hence can't create internal representations of values of clocked voltage levels, etc. And if there are no internal representations of such values, then instances of the 2-term internal relation can't be established. Only meanings, or components of meanings, exist.

The point being that Searle says the things computers process have an extrinsic meaning but no intrinsic meaning. And the Chinese room thought experiment and CRA are based on this idea. That is, the idea that the only pace a computer could conceivably get a semantics from, or the only semantics that exist, are the extrinsic meanings of the shapes of the things the machine allegedly processes. That the only semantics that exist are inside the brains of observing humans. Quite frankly, this seems a very bizarre position. Where does human inner semantics come from?

Actually, the things computers process have neither an extrinsic nor an intrinsic meaning. Neither, it seems clear, do the electrical and chemical things organic brains process. So inner semantics exist. Maybe a computer could get one in a relevantly similar way to how humans do.

HUMAN SYMBOLS COMPARED TO MACHINE "SYMBOLS"

So to compare, in the human case symbols are external tokenized interpretable shapes. In the machine case they are internal meaningless tokenized values of imperceptible properties.

Imperceptible to humans, that is. But can the machine perceive them? The machine "symbols" for biological reasons might be devoid of meaning to humans, but do or could the values of the properties of machine data units have meanings to the machine?

AGI seeks human-like general intelligence in an electronic device. The human perceives external symbols with eyes. If the machine is human-like then it similarly will be unable to perceive the things computers process.

Humans cannot perceive their own neural pulses. They lack the sensory apparatus to causally react to (detect) microscopic electrical pulses propagating inside cellular extensions. So the answer is that the things computers process have no meanings either to the human or to the machine.

THE INTRINSIC MACHINE IS NOT A COMPUTER

According to Zenon Pylyshyn,

> "*I spoke of interpretation, or* designation*, as being a key relationship into which symbols in a computer enter. ... To count as a computation [the computer] must contain symbols that are interpreted. In other words, the symbols must represent numbers, letters, or words, etc. (the slogan here is: 'no computation without representation')*".

Further,

> "*[This designation] is not wired into the machine (for which all symbol tokens are meaningless, except that certain ones can be made to cause the execution of primitive operations). The designation is provided by a person who takes the symbols to be about something – that is, the person gives the symbols a semantic interpretation*".

And,

> "*Of course, the task of ... communicating to others what representation of internal symbols a user has in mind is greatly simplified by labeling keys with letters and numbers and wiring certain states to cause letters and numbers to be printed. ... We have already considered the vexing question of whether [an external semantics] is a matter of necessity or whether machines can have 'original' semantics*"[29].

Pylyshyn says computers process symbols, then explains that to the machine itself these things are parenthetically meaningless (but doesn't deny that computers intrinsically compute). This seems an unhelpful conflation. I think it's a clarification to restrict the concept of symbol to interpretable values of perceptible properties, consistent with Turing's 1936 explanation of human computing – and Pylyshyn's explanation of human computing.

I've tried to argue that in order to adequately address "*the vexing question of whether ... machines can have 'original' semantics*", the device called a "computer" must be viewed as a machine that intrinsically never performs, and could never perform, computations.

---

[29] Zenon Pylyshyn (1984), *Computation and Cognition: Towards a Foundation for Cognitive Science*, MIT Press, pages 62-63.

Although the things computers internally process are strictly meaningless both to humans and to the machine, attachments to the machine, as Pylyshyn notes, do have interpretable shapes deposited on them – typically on keys of the keyboard, and sprayed onto or fused into sheets of paper by printers. The issue being that the shapes printed on exposed surfaces of the input attachment and on sheets of paper emitted by the output attachment, these shapes are deposited on these exposed surfaces specifically and only to facilitate the human use of the machine.

Given the intrinsic meaninglessness of inner machine "symbols", this leads to an important conclusion. If Pylyshyn's statement, "*no computation without representation*" is accepted, then a computer might be performing computations to a human who sees and interprets the shapes of the encrusted symbols printed on or created by attachments. But to the machine itself, the intrinsic device, internally, it never computes.

And to see how a machine could think, AI needs to understand the intrinsic device. AI (horror of all horrors) should realize that the device it is calling a "computer" is given that name only because of its human use. The device in and of itself is a non-computer. Turing was wrong. To AI, the core issue is non-computing machinery and intelligence. And happily, AI already has a non-computer. It's the device for historical reasons of human use, called a "computer".

SIMPLER IDEA

So if the compound idea of computing is detrimental to understanding the essence of the machine, what is a simpler and better idea?

My own view is that all knowledge gained thought sensory experience is reducible to instances of the relation of temporal contiguity. Weirdly, I very strongly believe this, that is, that certain reactions to instances of temporal contiguity between data units emitted by sensors are the base algorithms of perception and what might be called perceptual knowledge (so seemingly excluding skill knowledge which also concerns effectors).

Reacting to instances of togetherness in time implies benefits. Respective algorithms are very small, very simple, and hence very fast. This is because the properties of the terms of the relationship (properties of the objects related) are irrelevant to the fact of temporal adjacency, so the algorithms lack conditionals indexed on values of properties of the objects. There is no different reaction to different shapes or values of any other property (matching and counting are allowed).

Another seeming benefit of the idea of reacting to instances of temporal contiguity is that respective processing is non-teleological. Data units arrive from sensors. They have values of properties, but the machine is not programmed to react to different values in different ways. (But the values are still there.) Repetition is counted, then what happens internally depends on the frequency count and property value counted, and the frequency depends on which values of the properties arrive and their quantities, and which values arrive and their quantities depend on the sensed environment (which of the values defined in the "symbol set" designed into the sensor are emitted by the sensor). Repetition count is simply a standard feature of many associative ideas of learning including James Gibson's idea of ecological perception including detection of invariance, and feature extraction by artificial neural networks including in deep learning.

A further possible benefit is that data units emitted by sensors, these are the terms of instances of the relation of temporal contiguity between the units emitted by sensors, are

substance extended in 3-dimensional space, and these units are processed according to their togetherness in time. One would expect any primitive essence of knowledge to embody aspects of both space and time.

Temporal and spatial togetherness are primitive associative relations. Explaining in detail how they accounts for intelligence, listing and explaining the algorithms, is a further matter. But having abandoned the idea of reaction to shape, to computation, and having embraced the idea of reaction to togetherness in time, analysis reveals that computers can record the temporal relation between incoming sensory data units as structure, and can build structure, for example using pointers, and populate it with data units.

For tree structures of connections and nodes, data units would usually be leaf nodes, and a complete such structure, which would be a forest of trees, would be mostly recorded instances of the relationship of temporal contiguity between data units emitted by sensors[30].

CONCLUSION

Turing in 1950 said he could only see a short distance ahead. He also then said he could see plenty that needs to be done. A few years later, his life was tragically cut short in mysterious circumstances. It was what researchers didn't see, and what is usually so difficult to see, that stalled progress. AI didn't question the fundamental conceptual understanding of its machine. It uncritically accepted the concepts provided to it by computer science for understanding the human use of the device, computational concepts, when what AI needed to understand was the machine in and of itself.

AI was too quick to claim a plinth in the Hallowed Halls of Science. The adopted computer science concept of machine computation led it into error. The project supposed that the machine processes symbols, a fiction that makes human use of the device easier for humans to understand. In fact, symbols are encrusted only on exposed surfaces of machine attachments, put there to facilitate human use. They are completely irrelevant to the intrinsic device.

John Searle's CRA accepts AI's flawed computer science conception of the computer. This makes the CRA unsound. Searle is right when he concludes that no computation could ever think. But the devices for historical reasons of human use called "computers", intrinsically don't perform computations.

In fact the things computers internally react to have neither an intrinsic nor an extrinsic semantics. Which is not entirely problematic, since it seems quite clear that neither do the inner units human brains react to.

In October 2012, Oxford physicist David Deutsch in The Guardian newspaper wrote, "*Philosophy will be the key to unlocking artificial intelligence*". Philosopher John Searle's Chinese Room Argument is unsuccessful, but provides an analysis that clarifies AI's conceptual terrain. Although the CRA doesn't identify the flaws in these concepts, the precision the CRA brings to bear makes the flaws less difficult to hunt down.

Once the idea of computation is abandoned, simpler ideas can present a more fundamental and accurate view of the device for historical reasons of human use called a "computer". One simpler idea is that all knowledge gained through sensory experience is reducible to instances of the relation of temporal contiguity. It's not inconceivable that a computer realizing this principle could develop an inner semantics.

[30] An analysis of some ideas about this is in "Learning Chinese" available online at associative-ai.com.