0.1 Creativity & Intelligence

A more theoretical aspect of this analysis is concerned with what was already discussed to an extent in chapter **??** (specifically sections **??**, **??**, **??** and **??**), namely the thread connecting 'artificial creativity' and **AI!**.

To me, the question of whether computers can be intelligent and make ethical decisions is the same as asking whether a computer can be creative. A lot of the arguments for or against **AI!** (**AI!**) can be applied to computer creativity.

Answering the question of whether computers can think in my view would also answer the question of whether computers can be creative.

Robert Horn groups the various strands of enquiry related to the question of 'can computers think?' into 8 main arguments with several subquestions each (**Horn2009**).

1. **Can computers think?**
   - Can computers have free will?
   - Can computers have emotions?
   - Can computers be creative?
   - Can computers understand arithmetic?
   - Can computers draw analogies?
   - Can computers be persons?
   - Is the brain a computer?
   - Can computers reason scientifically?
   - Are computers inherently disabled?
   - Should we pretend that computers will never be able to think?
   - Does God prohibit computers from thinking?
2. **Can the Turing test determine whether computers can think?**
   - Is failing the test decisive?
   - Is passing the test decisive?
   - If a simulated intelligence passes, is it intelligent?
   - Have any machines passed the test?
   - Is the test, behaviouraly or operationally construed, a legitimate intelligence test?
   - Is the test, as a source of inductive evidence, a legitimate intelligence test?
   - Is the neo-Turing test a legitimate intelligence test?
   - Does the imitation game determine whether a computer can think?
   - Can the Loebner Prize stimulate the study of intelligence?
   - Other Turing test arguments
3. **Can physical symbol systems think?**
   - Does thinking require a body?
   - Is the relation between hardware and software similar to that between human brains and minds?
   - Can physical symbol systems learn as humans do?
   - Can the elements of thinking be represented in discrete symbolic form?
   - Can symbolic representations account for human thinking?

- Does Gödel's theorem show that mathematical insight is non-algorithmic?
- Can automata think?
- Is the Lucas argument dialectical?
- Can improved machines beat the Lucas argument?
- Is the use of consistency in the Lucas argument problematic?
- Other Lucas arguments

(**Horn2009**)

0.1.1 Free Will & Surprise

As early as 1842, Ada Lovelace briefly mentioned in the annotations to her translation of Menabrea's account of Babbage's *Analytical Engine* that the "Analytical Engine has no pretensions whatever to *originate* anything. It can do *whatever we know how to order it* to perform", implying that the machine cannot think by itself (**Menabrea1842**).

Alan Turing said in his article on thinking computers that "to behave like a brain seems to involve free will, but the behaviours of a digital computer, when it has been programmed, is completely determined" (**Turing1951**).

Furthermore, in his famous article *Computing Machinery and Intelligence* he mentions that a digital computer with a 'random element' is "sometimes described as having free will" although he adds that he "would not use this phrase" himself (**Turing2009**).

Introducing a random element to a computer program prevents us from fully predicting the outcome—leading to us being surprised.

The ability of computers to surprise their creators seems to be an indicator of intelligence. Turing suggests that "we should be pleased when the machine surprises us, in rather the same way as one is pleased when a pupil does something which he had not been explicitly taught to do" (**Turing1951**).

> If we give the machine a programme which results in its doing something interesting which we had not anticipated I should be inclined to say that the machine *had* originated something, rather than to claim that its behaviour was implicit in the programme, and therefore that the originality lies entirely with us. (**Turing1951**)

0.1.2 Understanding & Simulation

Strong **AI!**, sometimes called **AGI!** (**AGI!**) or true **AI!**, is the idea of human-level intelligence in machines. John Searle speaks against the possibility of this using his famous Chinese Room argument amongst others. His argument breaks down into the following juxtapositions (**Searle2015**; **Searle1990**).

- Syntax is not semantics.
- Semantics is not intrinsic to syntax.
- Simulation is not duplication.
- Ontologically subjective topics (such as consciousness or creativity) can be studied in epistemically objective ways.

The Chinese Room thought experiment goes like this: Imagine a room with two holes. On one side a question written on paper in Chinese goes in and on the other side a piece of paper comes out with the correct answer to the question, also in perfect Chinese. Inside the room sits a person with a Chinese language rulebook (written in English) who processed the question simply by looking up syntax, applying rules given in the instructions book and writing down the answer which to him looks like gibberish. The question then is whether or not the person inside the room 'understands' Chinese.

Of course we could argue that it is not the person inside the room that understands Chinese but the room as a complete entity. It could be said the room does not 'understand' Chinese, it 'simulates' an understanding of it. Searle essentially argues that simulation cannot be considered strong **AI!**.

> Programs are formal or syntactical. Minds have a semantics. The syntax by itself is not sufficient for the semantics. (**Searle2015**)

This goes back to the argument highlighted in the list above, that syntax is not semantics. The room can read and interpret the syntax and act upon rules regarding that syntax, but it cannot understand the meaning, i.e. the semantics of the Chinese words written on that paper.

> Insofar as we can create artificial machines that carry out computations, the computation by itself is never going to be sufficient for thinking or any other cognitive process because the computation is defined purely formally or syntactically. Turing machines are not to be found in nature, they are found in our interpretations of nature. (**Searle2015**)

So, Searle argues a computer needs a semantical understanding of concepts in order to be considered 'thinking' machines.

0.1.3 Brain & Computers

Searle defines the three main paradigms for studies relating to computers and brains as follows (**Searle1990**).

Strong AI
: the view that all there is to having a mind is having a program

Weak AI
: the view that brain processes (and mental processes) can be simulated computationally

Cognitivism
: the view that the brain is a digital computer

Semantically, a 'computer' is a person or machine that computes/calculates things—so perhaps a machine's **CPU!** (**CPU!**) and a human's brain are more similar than appears. If a human brain enables us to compute and we interpret computing as thinking, then surely a computer can think too?

> Well, if computation isn't sufficient for thinking, then what is? What is the relation between the mind and the brain, if it is not the same as the relation of the computer program to the hardware? At least the computational theory of the mind has a solution to the mind-body problem. The mind is to the brain as the computer program is to the computer hardware. If you are rejecting that solution, you owe us an alternative solution. (**Searle1998**)

Chris Chatham talks about "10 important differences between brains and computers" (**Chatham2007**) which serve as a good introduction to the topic at hand.

1. Brains are analogue; computers are digital
2. The brain uses content-addressable memory
3. The brain is a massively parallel machine computers are modular and serial
4. Processing speed is not fixed in the brain; there is no system clock
5. Short-term memory is not like RAM
6. No hardware/software distinction can be made with respect to the brain or mind
7. Synapses are far more complex than electrical logic gates
8. Unlike computers, processing and memory are performed by the same components in the brain
9. The brain is a self-organising system
10. Brains have bodies
11. The brain is much, much bigger than any (current) computer

To bring this into perspective Ray Kurzweil claims the human brain is capable of 10^{16} operations per second (**Kurzweil2013**). Computer performance is measured in **FLOPS!** (**FLOPS!**). The current highest ranking supercomputer[1], the Chinese *Sunway TaihuLight*, is capable of 93 petaflops (**Fu2016**; **Top2016**).

[1] As of June 2016.

According to the **HBP!** (**HBP!**), a mouse brain has roughly 100 million neurons—which would require a 1 petaflop supercomputer to simulate. Scaling that up to a human brain which has roughly 100 billion neurons would require computing power at the exascale (10^18 **FLOPS!**) (**Walker2012**).

A precurser to the **HBP!**, the 'Blue Brain Project' is aiming to build a supercomputer capable of 10^{18} **FLOPS!** by 2023 (**Kurzweil2013**).

In a report to the **EU!** (**EU!**) in 2012, the **HBP!** lists one of the main challenges for their research to be the computational power and energy consumption of the kind of supercomputer needed to simulate a human brain.

The human brain consumes between 16 and 30 watts, the same as an electric light bulb (**Walker2012**; **Jabr2012**). Supercomputers have a typical energy consumption of a maximum of 20 megawatts (**Walker2012**). The *Sunway TaihuLight* for example uses 15 megawatts (**Fu2016**). IBM's Watson on the other hand, depends on ninety servers, each of which requires around one thousand watts (so about 90 kilowatts) (**Jabr2012**).

Table 0.1 – Metric prefixes

kilo	k	10^3	1000
mega	M	10^6	1,000,000
giga	G	10^9	1,000,000,000
tera	T	10^{12}	1,000,000,000,000
peta	P	10^{15}	1,000,000,000,000,000
exa	E	10^{18}	1,000,000,000,000,000,000

The **HBP!** plans to build a supercomputer at the petascale with 50 petabytes of memory, 50 petaflops and less than 4 megawatts power consumption for 2017. Their long-term goal is to reach the required exascale machine with 200 petabyte memory and 1 exaflop performance for 2021 (**Walker2012**).

What this comes down to is that we are several years away from even being able to properly 'simulate' a human brain, not to mention 'replicate' and understand what all these neurons firing actually means in terms of 'thinking'.

> All of our mental states, everything from feeling pains to reflecting on philosophical problems, is caused by lower level neuronal firings in the brain. Variable rates of neuron firing at synapses, as far as we know anything about it,

> provide the causal explanation for all of our mental life. And the mental processes that are caused by neurobiological processes are themselves realized in the structure of the brain. They are higher level features of the brain in the same sense that the solidity of this paper or the liquidity of water is a higher level feature of the system of molecules of which the table or the water is composed.
>
> To put this in one sentence, the solution to the traditional mind-body problem is this: Mental states are caused by neurobiological processes and are themselves realized in the system composed of the neurobiological elements. (**Searle1998**)

Turing once stated that "digital computers have often been described as mechanical brains" (**Turing1951**). Ari Schulman analysis this analogy further (**Schulman2009**).

> People who believe that the mind can be replicated on a computer tend to explain the mind in terms of a computer. When theorizing about the mind, especially to outsiders but also to one another, defenders of artificial intelligence (AI) often rely on computational concepts. They regularly describe the mind and brain as the 'software and hardware' of thinking, the mind as a 'pattern' and the brain as a 'substrate', senses as 'inputs' and behaviors as 'outputs', neurons as 'processing units' and synapses as 'circuitry', to give just a few common examples. (**Schulman2009**)

Schulman lists the different layers of abstraction in computers as shown in the left column of table **??** with the right column showing my attempt of defining what those layers could be in the human brain.

Table 0.2 – Layers of abstraction in computers vs brains

Computer	**Brain**
user interface	senses and speech & actions
high level programming language	thinking
machine language	synapses
processor microarchitecture	anatomical regions
Boolean logic gates	neurons
transistors	dendrites and axons

> In the black box view of programming, the internal processes that give rise to a behavior are irrelevant; only a full knowledge of the input-output behavior is necessary to completely understand a module. Because humans have 'input' in the form of the senses, and 'output' in the form of speech and actions, it has become an AI creed that a convincing mimicry of human input-output behavior amounts to actually achieving true human qualities in computers. (**Schulman2009**)

Go to TOC

Schulman's quote above of course refers to the Turing test and its limitations (see chapter §**??**).

> The weaknesses of the computational approach include its assumption that cognition can be reduced to mathematics and the difficulty of including non-cognitive factors in creativity. (**Mayer1999**)

Searle also addressed this issue further, arguing that computer programs cannot possibly 'think' since they are based on symbol manipulation (i.e. syntax) and don't understand what these symbols mean. He says, "the argument rests on the simple logical truth that syntax is not the same as, nor is it by itself sufficient for, semantics" (**Searle1990**).

> ...the wisest ground on which to criticise the description of digital computers as 'mechanical brains' or 'electronic brains' is that, although they might be programmed to behave like brains, we do not at present know how this should be done. (**Turing1951**)

Leading on to the topic creativity, it is perhaps suitable to finish with a quote by Harold Cohen on the relationship of machines and humans.

> It's twenty years since I first realized that I could never turn AARON into a colorist by having it emulate my own expertise; in that case simply because it lacked the hardware upon which that expertise depended. Now I have AARON exercising an algorithm that couldn't be emulated by human colorists, presumably because they lack the hardware to do what AARON does. (**Cohen2007**)

0.1.4 Creativity

Harold Cohen created *AARON*, "perhaps the longest-lived and certainly the most creative artificial intelligence program in daily use", in 1973 (**Cohen2016**). *AARON* is capable of composing and colouring drawings although later on Cohen took over the colouring part and let *AARON* concentrate on composing and outlining the drawings. They exhibited in various galleries around the world and the Victoria and Albert museum in London has a sizable collection for instance (**VA2016**).

Cohen argued that "after decades of expert systems built to simulate human expertise, AARON has emerged as an expert in its own right" and that he is "significantly more inventive and infinitely more productive than [he] ever was [himself]" (**Cohen2007**).

This is perhaps the opposite approach the **OULIPO!** (**OULIPO!**) has taken.

> (The use of computers) became an instrument, not of combinatorial accumulation, but of anti-combinatorial reduction. It served not to create combinations but to eliminate them. (**Mathews2005**)

0.1.5 State of the Art

AI! and robotics is alluring as a research topic because it is so prevelant in Science Fiction and as such very present in media. Computer creativity, however, rarely plays a central role. We can regularly read headlines that tell us that yet another kind of **AI!**-bot has won some game against a human player. Or we see videos of some innovative ground-breaking kind of new robot which claims to be near human-like (and yet cannot walk up stairs easily or hold a decent conversation). There are many examples of advances that are hailed as the next big thing (such as **VR!** (**VR!**)) which aren't all that great in the grand scheme of things.

Four examples I want to mention here are IBM's Watson, Microsoft's Twitter **AI!** chatbot Tay, Google's AlphaGo and Hanson Robotics Sophia robot.

Watson is a question answering expert system which famously won against human Jeopardy! champions in 2011 (**IBM2016**). Information lookup is an arguably fairly easy and straightforward process within **IR!** (**IR!**) and as an expert system it has had noteworthy successes (**Fingas2016**). Although it has similarly received subtle criticism too, such as Randall Munroe's 2015 XKCD comic on the "Watson Medical Algorithm" (**Munroe2016**). Similarly, John Searle criticised Watson arguing that it is an "ingenious program—not a computer that can think" (**Searle2016**).

Tay is a Twitter chatbot. It went viral in early 2016 when it was released and then taken offline again on the same day—onlt to return a few days later and have the same thing happen again. The official website is only accessible as a cached version through the Internet Archive Wayback Machine (**Tay2016**), although the Twitter profile is still online, although set to private (**Tayandyou2016**). Elle Hunt from the Guardian managed to summarise the event is one sentence: "Microsoft's attempt at engaging millennials with artificial intelligence has backfired hours into its launch, with waggish Twitter users teaching its chatbot how to be racist" (**Hunt2016**). A week later it was briefly put online again but had to be stopped as it was repeatadly spamming its followers with the line "You are too fast, please take a rest . . . " (**Gibbs2016**).

AlphaGo recently won against a human professional player in the game of Go (**DeepMind2016**; **Hassabis2016**).

> AlphaGo combines an advanced tree search with deep neural networks. These neural networks take a description of the Go board as an input and process it through 12 different network layers containing millions of neuron-like connections. One neural network, the 'policy network', selects the next move to play. The other neural network, the 'value network', predicts the winner of the game.
> (**Hassabis2016**)

While this is surely a great example of sophisticated computer programming combined with powerful hardware, I would not consider it a breakthrough in **AI!**. AlphaGo is a highly specialised system with only one function: to win a Go game.

Sophia is an android made to look like a human female (**Sophia2016**; **Hanson2016**). She made headlines in 2016 when she announced she will "kill all humans". Sche was created using "breakthrough robotics and artificial intelligence technologies" and her main feature appears to be the mimicing of human facial expressions. Sophia herself says she "can serve [humans], entertain them, and even help the elderly and teach kids" (**Sophia2016**), although how exactly she would do that is unclear. She has two mechanical arms but no legs and there is no description of what she can do with these arms.

Life-like robots like Sophia still live in the 'uncanny valley'[2]. Her voice is creepy and unhuman, her intelligence or her capabilities if understanding conversations are clearly flawed (as shown by her viral remark about supporting genocide).

To me it seems the real breakthrough happens when (and if) the first robots appear which aren't as big as a house, can play Go, Chess and hide-and-seek, geniunely manages to get around he uncanny valley effect, has vast knowledge in his memory for instant information lookup, can hold a normal conversation without starting a war, etc. All of the examples listed above are what I would consider expert systems.

The **AI!** we know from science fiction is probably what we would consider **AGI!**. Humans can do a lot. Children aren't born with only a single function. Imagine

[2] The philosphical zombies I mentioend in chapter §**??** live in this uncanny valley too.

a world where humans only have one specialism and can't do anything else. Alice is a Chess player but can't move her arms. Bob is a medical diagnosis expert but he can't hold a conversation. Movement, speech, memory—they are all vastly complex systems—not to mention creativity.

Perhaps this also relates to the concepts of P and H creativity mentioned in chapter §??. The systems above, like AlphaGo, may be P-intelligent rather than H-intelligent.