Tancredi Di Giovanni

OUT-OF-HARDWARE EXPERIENCE

software and consciousness

Thesis submitted to: the Department of Experimental Publishing,
Piet Zwart Institute, Willem de Kooning Academy,
in partial fulfilment of the requirements for the final examination for the degree of:
Master of Arts in Fine Art & Design: Experimental Publishing.
Adviser: Steve Rushton
Second Reader: Marloes de Valk
Word count: 9300

Table of Contents

Introduction

The modern theories on the relation between humans and machines are, to a great extent, a consequence of the scientific positivism developed in the Western world during the 19th century, which assumes that the subjective nature of human consciousness can be denied or reduced to its objective physical substance: the brain. This speculation models the relation between humans and machines on two axes:

[x] *It emphasises the relation between a loss of humanity and the rising of autonomous machines* (this view implicitly portrays software as the consciousness of the machine originated by a hardware).

[y] *It emphasises the relation between the social system and the technical system* (this view explains software as a cultural and social object that can be studied independently from the hardware).

In order to contain the exponential growth of the technical system and the disappearance of the individual, the institutional power becomes the only legitimate means of containment imposing regimes of engineered subjectification while enmeshing society with its technologies. Fostering visions of dystopic futures and lacking of early systematic critiques, these arguments silently underpin the technological and social developments of the 20th century, to reach daily life in our contemporary society.

In *Part 1*, unified under the umbrella term *machinic life* coined by John Johnston, this thesis looks back at the general history and theoretical results of attempting to build autonomous machines[1] *"[…] mirroring in purposeful action the behavior associated with organic life while also suggesting an altogether different form of "life" […]"* (Johnston, 2008). In opposition to the assumptions of machinic life, in *Part 2*, this thesis proposes a different approach informed by the new developments in the understanding of consciousness:

[z] It *emphasises the relation between subjective experience and the technical system***,** *unfolding a more clear understanding of both biological and artificial systems as part of an extended cognitive system.*

In this direction, I've found a particular resonance of my thoughts with the work of David Chalmers on consciousness, claiming for a paradigm shift in science to finally allow the study of subjective experience, *"when simple methods of explanation are ruled out, we need to investigate the alternatives. Given that reductive explanation fails, nonreductive explanation is the natural choice"* (Chalmers, 1995). Thomas Metzinger, who, through the study of altered states of consciousness and psychiatric syndromes, is one of the few to propose an appealing alternative (reductionist) model of consciousness capable of explaining the nature of the self, *"If we pay more attention to the wealth and the depth of conscious experience, if we are not afraid to take consciousness seriously in all of its subtle variations and borderline cases, then we may discover exactly those conceptual insights we need for the big picture"* (Metzinger, 2009). And Katherine N. Hayles, who, starting from a social perspective and criticizing consciousness, links machines and biological systems

1 Not meant to be exhaustive, this historical account of *machinic life (*which comprehend *Cybernetics*, *symbolic* and *sub-symbolic AI* and a brief mention of *Alife)* provides a sporadic overview of a much more interesting and articulated story that is still in progress and which detailed account can be found elsewhere.

extending cognition into the body and the environment, *“Although technical cognition is often compared with the operations of consciousness [...], the processes performed by human nonconscious cognition form a much closer analogue”* (Hayles, 2017). And Matteo Pasquinelli, for his studies on *machinic intelligence* in general, but always source of great inspiration.

Instead of addressing its technical and cultural aspects directly, the result of these discourses reveals a new primary condition of software pointing toward the subjective experience of new phenomenal worlds that can be built and sustained in collaboration with an external artificial form of cognition. Knowledge, from this point of view, is not inaccessible to the individual level, given by science and institutionalized through society, but is a necessary process made through the construction of worlds, simulating scientific truths or creating useful fictions, but still validating the subject as the designer (or hacker) of its own experience.

“But why should I repeat the whole story? At last we came to the kingly art, and enquired whether that gave and caused happiness, and then we got into a labyrinth, and when we thought we were at the end, came out again at the beginning, having still to seek as much as ever.”

— Plato, Euthydemus

PART 1

“Then, just as the frightened technicians felt they could hold their breath no longer, there was a sudden springing to life of the teletype attached to that portion of Multivac. Five words were printed: INSUFFICIENT DATA FOR MEANINGFUL ANSWER.”

— Isaac Asimov, *The last Question*

The hard problem of consciousness

Through the standard scientific method the challenge of explaining the mind has been mostly addressed by disassembling it in its *"functional, dynamical and structural properties"* (Weisberg, 2012). Consciousness has been described as cognition, thought, knowledge, intelligence, self-awareness, agency and so on, with the assumption that explaining the physical brain would resolve the mystery of the mind. From this perspective, our brain works as a complex mechanism that eventually triggers some sort of behavior. Consciousness is identified with a series of physical processes happening in the cerebral matter *(reductionism)* and determining our experience of having a body, thinking, and feeling. This view has been able to explain many unknown elements of what happens in our minds.

In 1995 the philosopher of mind David Chalmers published his article *Facing up to the problem of consciousness*, where he pointed out that the objective scientific explanation of the brain can solve only an easy problem. If we want to fully explain the mystery of the mind, instead, we have to face up the *hard problem* of consciousness: *How do physical processes in the brain give rise to the subjective experience of the mind and of the world? Why is there a subjective, first-person, experience of having a particular kind of brain?* (Nagel, 1974)

Explaining the brain as an objective mechanism is a relatively easy problem that eventually, in time, could be solved. But a complete understanding of consciousness and its subjective experience is a hard problem that scientific objectivity cannot access directly. Instead, scientists have to develop new methodologies and eventually non-reductive models, considering that a hard problem exists — *How is it possible that such a thing as the subjective experience of being 'me, here, now' takes place in the brain?*

Echoing the *mind-body problem* initiated by Descartes in the 17th century, subjective experience, also called *phenomenal consciousness* (Block, 2002), underlies any attempt to investigate the nature of our mind. It challenges the physicalist ontology of the scientific method showing the unbridgeable *explanatory gap* (Levine, 2009) between the latter's dogmatic view and a full understanding of consciousness. This produces the necessity of a paradigm shift allowing new alternative scientific methods to embrace the challenge of investigating phenomenal consciousness for example the *neurophenomenology* proposed by Francisco Varela (1996).

The reactions to Chalmers' paper ranges between a total denial of the issue to panpsychist positions, with some isolated cases of mysterianism advocating the impossibility to solve such a mystery (Weisberg, 2012). In any case, the last thirty years have seen exponential growth in multidisciplinary researches facing the hard problem with a constant struggle to build the blocks of a science of consciousness finally accepted as a valid field of study (Metzinger, 2009).

Hidden for ages behind the ambiguous religious beliefs in the *soul* and the immediacy of empirical evidence on which science is based, phenomenal consciousness is now at the very first stages of a proper scientific unfolding of its contents (Metzinger, 2009). Thanks to a renovated view of science and its methods, subjective experience is starting to be resized to its effective dimension filling the gaps in the understanding of ourselves and the world. But before we explore in depth the contents of phenomenal consciousness and their implications in understanding software, it is essential to shift our attention toward the evolution of new kinds of machines in technical systems. Fostering the division between hardware and software and leading part of the scientific community to acknowledge the limits of their own practices, the understanding of the autonomous machine is the fundamental step in actualizing and testing the scientific modeling of the mind.

Machinic life and its discontent - I

In *The Allure of the Machinic Life*, John Johnston (2008) attempts to organize the contemporary discourse on machines under a single framework that he calls *machinic life*[2].

> *"By machinic life I mean the forms of nascent life that have been made to emerge in and through technical interactions in human-constructed environments. Thus the webs of connection that sustain machinic life are material (or virtual) but not directly of the natural world."* (Johnston, 2008)

Machinic life, unlike earlier mechanical forms, has a capacity to alter itself and to respond dynamically to changing situations.Implying the whole attempt to produce life and its processes out of artificial hardware and software, the definition of machinic life allows us to re-consider the different experiences of the last century under the common goal of building autonomous machines and to understand their theoretical backgrounds and assumptions as a continuum.

Subsumed in the concept of *techné*[3], the mythological intuition of technology already shows the main paths of the contemporary discourse of machinic life. In fact, in the *myth of Talos* and in *Daedalus' labyrinth*, we can find the first life-like automaton and the first architectural design reflecting the complexity of existence and outsourcing thought from human dominion. However, only in the 19th century, with the new technological discoveries and scientific positivism[4], did scientists start building the bearing structures of what would become the two main fields of machinic life of the 20th century: *Cybernetics* and *Artificial Intelligence* (AI).

On one side, this process begins with the improvement of the steam engine with Sadi Carnot's *thermodynamics* (1824) joined with the *debate on the origin of life* opposing the theory of evolution to the religious belief in Creationism. Unleashed from the religious teleology (purpose) imposed by God's intelligent design and consigned to the random chance of natural selection introduced by Charles Darwin and Alfred Wallace's *evolutionary biology (1859)*, human existence was losing any perspective of independent agency (Rushton, 2019). In 1858, Wallace wrote a letter to Darwin comparing the evolutionary process with the vapor engine auto-regulatory system, or feedback loop, later studied by J. C. Maxwell's *control theory* (1868).

> *"The principle of this process* [natural selection] *is exactly like that of the centrifugal governor of the steam engine, which checks and corrects any irregularities almost before they become evident" (Wallace, 1858)*

With this same conclusions, Samuel Butler, speculating on the evolution of machines in texts such as *Darwin Amongst the Machines* (1872) and *Erewhon* (1879), reintroduced the idea of teleology in the concept of adaptation**,** developing a framework where machines are capable to evolve and reproduce exactly as biological organisms (Rushton, 2019). Wallace and Butler's speculative theories anticipated a new understanding of the machine that would be actualized only through the advancement of control theory in communication, guns automation, and biology occurred during World War II (Johnston, 2008). In 1946, many scientists, before working for military projects, gathered to model the auto-regulatory system of the body and simulate it in

2 Developed by Deleuze and Guattari's *machinic philosohpy* (1972, 1980) (Johnston, 2008), the term *machinic* stakes *"the existence of processes that act on an initial set of merely coexisting, heterogeneous elements, and cause them to come together and consolidate into a novel entity." (DeLanda, 1997).*

3 *"An art, skill, or craft; a technique, principle, or method by which something is achieved or created."* (Oxford Dictionary)

4 Formulated by Auguste Comte in the early 19th century, *positivism* rejects subjective experience because it is not verifiable by empirical evidence.

autonomous robots giving rise to *Cybernetics* defined as *Control and Communication in the Animal and the Machine* (Wiener, 1948).

Parallel to these developments, the study of mathematics and logic, along with the revolution of the *Jacquard's loom* (1804), led to the invention of the first *general-purpose computer* and the translation of elementary logic into binary algebra. Charles Babbage and Ada Lovelace's effort to design and program the *analytical engine* (1837), together with *Boolean logic* (1854), started a new era of computation in which mental labor was not anymore an exclusive prerogative of humans but could be performed by an economy of machinery. Demonstrated for computable numbers by Alan Turing and Alonzo Church's model of *computation theory* in 1936, the idea of formalizing thought in an instrumental set of rules (*algorithm*) can be traced back to Plato[5] and Leibniz[6] (Dreyfus, 1972). The *Church-Turing thesis*, together with Von Neumann's *computers architecture* (1945) and Shannon's *information theory* (1948) influenced by cybernetics, sign the birth of the digital computer making possible the beginning of *Artificial Intelligence* (AI) in 1956 (Russell&Norvig 2003).

If the classical world had the intuition of the sentient machine, and the modern the realization of its possibility, it is only with the practical experience of cybernetics and AI that the contemporary discourse of machinic life can be formulated. Nonetheless, this dual nature of contemporary discourse embodies the convergence of different theories in biological, mechanical and computational systems within a multidisciplinary approach, driven by complexity and information. Furthermore, as we will see in the next chapter, the limits of machinic life in understanding and building working models of the mind can already be seen in how cybernetics and AI equate the human nature with the nature of the machine leading to the distinction between hardware and software.

Machinic Life and Its Discontent - II

Consolidating at the *Macy conference* occurred in 1946 in New York City headed by Norbert Wiener, Arturo Rosenbleuth (1943) and Warren McCulloch (1943)[7] works, Cybernetics has been the first framework capable of generating a working theory of machines (Johnston, 2008). Its influence spreads in different disciplines such as sociology, psychology, ecology, economics as well as in popular culture (*cyberculture*). The prefix *cyber-*, in fact, will become emblematic of a new understanding of the human condition as profoundly intertwined with machines. Supported by statistical information theory, experimental psychology, behaviorism and control theory, Norbert Wiener (1948) saw in the process of adaptation of the body first described as *homeostasis* by Walter Cannon (1936), the possibility to simulate the same mechanism in autonomous artificial organisms. Transforming life in a complex adaptive system pairing an organism with its environment through feedback loops, this position conceptually leads to the dissolution of boundaries between natural and artificial, humans and machines, bodies and minds. Human beings and machines become cybernetic subjects, in a world where nature and life are no longer a matter of organic and inorganic substance but of structural complexity

---

5 *"I want to know what is characteristic of piety which makes all actions pious [...] that I may have it to turn to, and to use as a standard whereby to judge your actions and those of other men."* (as cited in Dreyfus, 1972)

6 *"Once the characteristic numbers are established for most concepts, mankind will then possess a new instrument which will enhance the capabilities of the mind to far greater extent than optical instruments strengthen the eyes, and will supersede the microscope and telescope to the same extent that reason is superior to eyesight."* (as cited in Dreyfus, 1972)

7 His seminal work is particularly relevant as a first comparison between the brain and the digital information processing anticipating both *computationalism* and *connectionism* (Dreyfus, 1972, Rescorla, 2020).

(Johnston, 2008). The implications of this view broke the boundaries of human identity, leading theorists to talk about *post-humanism* and explore new realms of control and speculations on the nature of machine simulation (Hayles, 1999).

Despite the variety of subfields developed by Cybernetics[8], the parallel advent of the digital computer obscured most of its paths for decades (Cariani, 2010). The focus of researchers and national funding shifted into the framework of Artificial Intelligence (AI). This new focus on intelligence, to which consciousness is allegedly a feature, was made possible by establishing a strict relation between the mind reduced to the brain and the digital computer. In fact, another revolution was taking place in the field of psychology. The incapacity of *behaviorism,* which looks at psychological processes as a matter of input and output, to include mental processes in understanding humans and animals, was opening the doors to the *cognitive revolution*. The mind, intended as the cradle of cognitive processes, was compared with digital computer's information processing, making it possible to test psychological theories and simulate the behavior of mental processes in the artificial brain[9] (Miller, 2003). Furthermore, this approach permits to extend the *mind-body dualism* in the machine as software and hardware. In contrast to cybernetics, promoting auto-regulation in biological and artificial organisms as *embodied knowledge* acquired through experience (Johnston, 2008)[10], AI and its subtending computers philosophy fosters the division between hardware and software, abstracting information processing from its physical ground and leading to the consequent obscuration of hardware through software (Kittler, 1992).

Before AI was officially born, in 1950 Alan Turing published his article *Computing machinery and intelligence*, where he designed the *imitation game* best known as the *Turing test*. The computational power of the computer was identified with the act of thinking understood as intelligence.

> *"The reader must accept it as a fact that digital computers can be constructed, and indeed have been constructed, according to the principles we have described, and that they can in fact, mimic the actions of a human computer very closely." (Turing, 1950)*

Because the phrasing of the problem as "c*an machines think?"* can have ambiguous results, to allow computer scientists to explore the possibility of creating intelligent machines, Turing reversed the question into a behavioral test — *Can we say a machine is thinking when imitating a human so well that s/he thinks s/he is talking to another human?* — If you can't recognize that your interlocutor is a machine, it doesn't matter if it is actually thinking, because in any case, the result would be the same, a human-level communication. Thinking and mimicking thinking become equivalent, allowing machines to be called intelligent. In his text, Turing dismisses the argument of phenomenal consciousness and the actual presence of subjective experience by sustaining that such a problem does not necessarily need to be solved before being able to answer his question. Indeed, the Turing test suggests more than a simple game. It is signaling the beginning of a new inquiry on the theoretical and practical possibility of building "*real*" intelligent machines while indicating some possible directions[11] to build a machine capable of passing his test. (Dreyfus, 1972, Rescorla, 2020).

Riding the new wave of the cognitive revolution and embracing the cybernetic comparison between humans and machines, a group of polymaths began to meet in 1956 at Dartmouth College, the birthplace of AI. Developed by Allen Newell and Herbert A. Simon the *Logic Theorist* was the first working program exploring

8 *"[...] self-organizing systems, neural networks and adaptive machines, evolutionary programming, biological computation, and bionics."* (Cariani, 2010)

9 Hubert Dreyfus (1972) refers respectively as cognitive simulation (CS) and artificial intelligence (AI) *"in a narrower sense"*.

10 Johnston (2008) explicitly opposed its view to the argument of Hayles (1999) *"that cybernetics […] effected a 'disembodiement' of information by defining it independently from its material substrate"*.

11 Natural language processing, problem-solving, chess-playing, the child program idea, and genetic algorithms.

the automation of reasoning through its formalization and manipulation in a symbolic system. Called *Symbolic AI*, this approach will become the workhorse leading the expected escalation from programs limited to reproduce only a narrow task, *narrow artificial intelligence* (NAI), to programs capable of doing any task, *artificial general intelligence* (AGI), and achieve the level of human intelligence, *human-level artificial intelligence* (HLAI), as prospected by Turing (Russell&Norvig 2003). Exactly because of this overstated goal and expectations, the fathers of AI[12] will be remembered as the enthusiastic researchers drawn in a spiral of premature predictions and hyperbolic claims (Dreyfus, 1972) that mostly failed, or, that they have not been achieved yet.

Infected by this early enthusiasm, psychologists and philosophers of science already struggling on the possible equation between the brain and the thinking machine, started to attempt a serious interpretation of the human mind based on the information processing of the new computational systems. This approach, called *computationalism*[13] led to several theories (Rescorla, 2020) such as: *The Computational Theory of Mind* (CTM) introduced in philosophy by Hilary Putnam (1967), basically understands the mind as a linear *input-processing-output* machine on the style of the computational model provided by Turing. Jerry Fodor's *Language of Thought Hypothesis* (LOTH) and its R*epresentational Theory of Mind* (RTM) (1975) claims thinking is only possible in a *language-like* structure to build thoughts at the top-level. The *Physical Symbol System Hypothesis* (PSSH) of A. Newell, H. Simon (1976) sees in the physical symbolic system everything needed to build a true intelligence. In popular culture as well, the same enthusiasm led to a new ideology of the machine with its climax in 1968 with the fictional character of *HAL 9000* of the movie *2001: A Space Odyssey* by Stanley Kubrick[14] (Dreyfus, 1972).

Despite the great enthusiasm and expectations, the idea that computers can do all the things a human can do, has been heavily criticized. Philosophers such as Hubert Dreyfus (1965) and Noam Chomsky (1968) started to highlight the problematic aspects of the computationalists theories of mind beginning a critical analysis of AI. They revealed the simplistic assumptions perpetuated by the unjustified hype and incapacity ofauto-criticism of major AI's researchers and showed the technical limitations of physical symbolic systems. Their inability to grasp the value of the context, essential in gaining knowledge and achieving common sense (Russell & Norvig 2003), and the impossibility to formalize all the aspects of intelligence, such as creativity and intuition (Dreyfus, 1972), were recognized as some of the principal boundaries in “decoding” the mind.

In the same direction, philosopher John Searle, criticizing the comparison of the human mind with computers in understanding things, developed a thought experiment called the *Chinese room*[15] (1980) arguing

12 In her account of the early days of AI, Pamela McCorduck (1979) recognizes Allen Newell, Herbert A. Simon, Marvin Minsky and John McCarthy as the former fathers of this discipline (McCorduck & Friedman, 2019).

13 This approach is often called functionalism in philosophy (Block, 2002) even though as a generalization (Rescrola, 2020).

14 HAL 9000 is depicted as a malevolent human-like artificial intelligence capable of feeling emotions, designed with the technical consultancy of Marvin Minsky (Dreyfus, 1972).

15 *“Suppose that I'm locked in a room and given a large batch of Chinese writing […]* [but] *to me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that 'formal' means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols […] from the point of view of somebody outside the room in which I am locked -- my ‘answers’ to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese.”* (Searle, 1980) In conclusion, a machine following a code, exactly as the person locked in the Chinese room, doesn’t “really” understand its inputs and outputs.

for an underlining distinction between a *strong AI* capable to really understanding, and a *weak AI* which just simulates understanding. Searle's argument raises the same issues of the aforementioned *hard problem of consciousness*, defining a threshold between the actual AI and the human mind. Other thought experiments, such as Jackson's *Mary's room*[16] (1986) touch the subjectivity of experience directly, which seems to resist all the efforts of the scientific community to reduce it to a machine and its weak computational intelligence.

Machinic Life and Its Discontent - III

The computational symbolic AI, believes that from a top-down approach you can engineer all the aspects of the mind in digital computers, including consciousness which is reduced to a mechanism of the brain. However, despite the early success (still limited when compared to the actual goals of HLAI), the wrong predictions, their conceptual limitations and the difficulty to find commercial applications for their work led to a series of failures resulting in two periods of recession between 1974-1980 and 1987-1993 best known as *AI winters* (Russell&Norvig, 2003). After these periods, the criticism moved toward the symbolic approach and the development of new research inspired by cybernetics, led AI researchers to understand intelligence and design life through different approaches called *sub-symbolics*.

Instead of an upstream representation of knowledge typical of the symbols manipulation, in 1943 cyberneticists Walter McCulloch (1946) were already looking closely at the architecture of the brain exploring the possibility to reproduce its networks of neurons in *artificial neural networks* (ANN). However, this system will become effective only from 1986 with Rumelhart, Hinton and McClelland’s *parallel distributed processing* (PDP) pairing multiple layers of ANNs and increasing drastically its capacity (Russell&Norvig, 2003). The use of ANN in AI explains intelligence from a bottom-up approach starting the paradigm called *connectionism*. These new AI systems are capable of learning and finding useful patterns by inspecting sets of data and reinforcing the connections between their neurons (Alpaydin, 2016). Thanks to the internet and the development of the last decade with the *deep learning* method, ANN's can now be fed with a large amount of data, increasing drastically their capacity to learn and producing a renovated hype in connectionism and AI.

Another relevant approach produced in AI by the late influence of cybernetics is the intelligent agent paradigm described by Stuart J. Russell and Peter Norvig in 2003. Reintroducing the discourse on complex systems, the concept of *rational agent* (borrowed from economics) becomes the way to refer to anything capable of interacting with an environment through sensors and actuators. AI systems developed in this perspective are capable of achieving a goal by keeping track of their environment, learning and improving their performance autonomously. In parallel with the developments in AI, cybernetics also highlighted the possibility of simulating biological evolution in software environments. Starting from the arrangement of simple operators called *cellular automata* (CA) on a grid and simple laws describing the possible interactions, neighbor cells

16 *“MARY is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about the environment […] If physicalism is true, she knows all there is to know. [...] It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning--she will not say ‘ho, hum.’ Hence physicalism is false.”* (Jackson, 1986) In conclusion, subjective experience can’t be reduced to a code and therefore ‘strong AI’ is not possible with symbolic AI.

starts to reproduce, die, and evolve, forming complex chaotic systems, stable loops and astonishing patterns that are impossible to predict *a priori* (Johnston, 2008).

These new ways of defining life and intelligence to correct the symbolic approach, are moving toward a deeper understanding of cognition that instead of being represented only as a symbolic system, it also lies on sub-symbolic level and instead of being a designed product, it is seen as part of evolutionary processes. However, despite these new developments, AI is encountering its boundaries. Life, like intelligence, generates from the interaction with an extremely complex and variegated environment, the *noisy physical world* which is made of radiations and electromagnetic phenomena, particles and wavelengths in continuous interaction. A chaotic world that neither contemporary computers' capabilities nor the internet's amount of data can simulate (Johnston, 2008). Furthermore, the companies relying on deep learning are looking into the problem of understanding why these learning systems make their choices. Their autonomous way of learning through layers of networked neurons creates nested black boxes extremely difficult to unpack rising a whole debate on discrimination and biases embedded in software. To escape from these limitations, scientists are now working on a more holistic understanding of intelligence which combines the sub-symbolic approach with the knowledge representation of symbolic AI (Marcus&Friedman, 2019). In robotics, the *situated AI* is rediscovering the necessity of having a body and taking robots outside of the labs to interact with the noisy physical world, hoping to find new ways to generate knowledge from direct experience instead of just simulating it in virtual environments (Russell&Norvig, 2003).

In the last 20 years, machinic life is starting to take seriously its critiques and reassessing the simulation of the *adaptive unconscious* and *embodied knowledge* typical of biological organisms (Kahneman&Friedman, 2011) as the possible link to produce the high-level intelligence typical of intuition, creativity and the spontaneous complexity of life. However, after almost 70 years from the first AI program, we are still surrounded by only *weak-and-narrow* AIs. On the one hand, part of the researchers reformulated the goal of building human-level systems, as well as ‘strong AI’, on less pretentious and more practical aims. On the other hand, despite its turbulent early history of unhappy claims and the slow growth of connectionism, the perspectives of engineering AGI and *strong AI* systems and populating the world with new forms of artificial life are growing faster, and with them, again, their premature claims.

Riding the regenerated hype made possible by the boom of deep learning, private institutions, such as MIT, tech entrepreneurs, such as Elon Musk[17], and many other researchers in AI's related fields, such as the futurist Ray Kurzweil[18], are repeating the same errors of the old fathers. They daydream a future-oriented, techno-utopianist world that directly reminds the morally dubious neo-liberal *Californian Ideology* (Barbrook&Cameron, 1995). This new goal, expressed by the MIT's spokesman Lex Friedman (2018), represent the *big picture* of AI which passes through the development of AGI and HLAI, to reach the *technological singularity* as described by the science fiction writer Vernor Vinge:

> *"a change comparable to the rise of human life on Earth. The precise cause of this change is the imminent creation by technology of entities with greater than human intelligence."* (Vinge, 1993)

Eventually, this black-boxed *super artificial intelligence* (SAI) will develop *artificial consciousness* on its own, emancipating and starting to “think for itself” becoming the form of life of the future (eventually helping, killing or snubbing human beings).

17 In particular, for some of the goals of its companies and its various claims on the advent of AGI (Musk&Friedman 2019).

18 In particular, for its previsions and its way of popularizing the advent of AGI in its books.

The AI's *big picture* sores the nervous system of popular culture. It creates misunderstandings on the actual state of affairs, clamorous expectations in the near futures and discouraging doubts on positive future perspectives. Furthermore, it provides a framework where the understanding of software and computers in general, relies on a matrix of abstractions obscuring their actual nature still made of man-made mechanisms and *weak AI* system relying on the labor of their producers and the interests of a capitalist market ready to exploit the fooled final user. AIs researchers, instead of disregarding the theoretical issues and technical limitations of their approaches, and follow the mainstream and commercially appealing *big picture* of AI, should convert their goals in developing a framework capable to confront with the problem raised by a study in depth of consciousness. *Machinic life* in general, should be re-framed to allow the study of the mind (and the body) with the primary goal to increase the scientific exploration of consciousness allowing a more complete understanding of nature. At this point in time, however, subjective experience still appear as what differentiate humans from machines, allowing us to imagine a present-oriented future where the *big picture* of AI is resized to its "weak" actuality and the focus is shifted to fix the natural and social problems that the mankind procrastinate with ignorance and presumption. This direction draws a more fruitful path that, through the understanding of consciousness, automatically leads to a better understanding of life, the world and the technical system allowing to design useful AIs with the awareness of its consequences on both the biological and artificial level. In the best case, if machinic life succeeds in engineering phenomenal consciousness and as professor Matteo Pasquinelli (2014) hopefully interprets the words of Turing, its result will be that of a new kind of alliance between the two forms of cognition.

Before proceeding with a detailed account of the characteristics of subjective experience, its similarities and differences with the computer and its relation with software, in the next chapter I will briefly introduce other approaches generated under the influence of machinic life. Instead of a focus on the autonomous machine, these frameworks re-frame the human-machine dichotomy developing the spaces in between these two extremes.

Beyond humans and machines

Confining human beings to a total subaltern level, yet fated to become futile, the intelligent and autonomous artificial organism conceived by cybernetics and AI implying an unsurpassable threshold between human and machine performance. However, in this power play configurations between natural and artificial agents, other possible worlds can be articulated. Worlds where humans and machines not only coexist but melt together achieving that level of close interaction between organisms known as *symbiosis* and leading to the paradigms of *intelligence augmentation* (IA) and *cyborg theory*.

On the one hand, and in parallel to AI, IA claims the possibility of augmenting human intelligence through technological means (Pasquinelli, 2014). Anticipated in 1945 with the prophetic words of Vannevar Bush, speculating on computers interface, and theoretically forged in cybernetics by W. Ross Ashby:

> *"[…] it seems to follow that intellectual power, like physical power, can be amplified. Let no one say that it cannot be done […]." (Ashby, 1956)*

Fostered by the visions of J.C.R. Licklider's *man-machine symbiosis* (1960) and Simon Ramo's *intellectronics* (1961) working in strict contact with the United States Department of Defense, the 60s have seen the consolidation of this promising paradigm in the development of *interactive computing* and the *user interface*. The work of Douglas Engelbart at the Aumentation Research Lab (ARL) and his political plan *bootstrapping human intelligence* (1962) automatically affecting society, will be remembered as the highest peak of IA before disappearing in the less politicized *human-computer interaction* (HCI) in the late 70s. Nowadays, a new frontier of *amplification, interaction and control* linking directly the brain with the computer is becoming possible (even though it is still at the very early stages). The *brain-computer interface* (BCI), gets us closer to that *"disturbing phenomena"* called *extrasensory perception "*[which] *seem to deny all our usual scientific ideas"* (Turing, 1950). The same BCI which Elon Musk's company *Neuralink* wants to develop, among other things, as a universal panacea to communicate with the artificial super intelligence of the dystopic near future (Musk&Friedman, 2019).

On the other hand, the disclosure of the cybernetic concept of life dissolves the human-machine dichotomy into an ecosystem of patch-worked organisms mixing together artificial and biological parts. This continuum, called *machinic phylum* by Deleuze and Guattari (1980) (Johnston, 2008), is the home of the *cyborg* (cybernetic organism) (Haraway, 1985), transforming its body in the playground where internal and external assemblages of parts, like implants different in their substance but communicating through feedback loops, coexists. The *cyborg theory* represents all the shades articulating the space between what is human and what is a machine. In this direction, Thomas Metzinger (2009) explains how *hybrid biorobotics* are another framework that is backing away from the pure artificial goal of AI and standard robotics, exploring the possibility of mixed species. The idea is that we can build artificial hardware running biological software as well as using artificial software to control biological hardware. If the first way tries to deploy patterns emerging in biological neural networks to run in artificial computers, the second finds its example in *RoboRoach* where the movements of a cockroach are controlled through an artificial implant sending electrical impulses to its nerves. This last approach, reconnecting to the BCI mentioned above, when it is used to directly stimulate the brain leads to what Metzinger (2009) calls *neuro-enhancement*, the artificial control of mental states (as the *neuro-* version of *psycho-pharmacology*). Due to the uncertainty of the assumption of machinic life sustaining that consciousness can be instantiated in a substance different than the biological *(biological assumption)* (Dreyfus, 1972), it seems that the control of the brain through artificial means could be an alternative way to achieve the synthesis of consciousness. Further technological developments in the field of BCI and in the design of non-neural hardware, will make possible to consider in which extent the biological assumption is effectively an assumption or an actual limit in building artificial consciousness.

All these different configurations and consequent understanding of the relation between the human and the machinic have a common denominator. The first step seems to be the much acclaimed *technological singularity*, intended (in less dystopic terms than the Vinge's version abovementioned) as *a particular moment in time in which there will be a drastic change in how we deal with technologies*. It could be the advent of AGI, HLAI or SAI. The construction of an affordable BCI or the rise of a cyborg society and the synthesis of artificial consciousness. But the final point, the farthest moment where theories conflate, is the *bio-digital fusion* that will follow the exponential growth of humans and machines and actualize the correspondence between the two systems.

> *"The stars and Galaxies died and snuffed out, and space grew black after ten trillion years of running down. One by one Man fused with AC* [Automatic Computer]*, each physical body losing its mental identity in a manner that was somehow not a loss but a gain. Man's last mind paused before fusion, looking over a space that included nothing but the dregs of one last dark star and nothing besides but incredibly thin matter, agitated randomly by the tag ends of heat wearing out, asymptotically, to the absolute zero." (Asimov, 1956)*

PART 2

“I am not advocating that we go back to an animistic way of thinking, but nevertheless, I would propose that we attempt to consider that in the machine, and at the machinic interface, there exists something that would not quite be of the order of the soul, human or animal, anima, but of the order of a proto-subjectivity. This means that there is a function of consistency in the machine, both a relationship to itself and a relationship to alterity. It is along these two axes that I shall endeavour to proceed.”

— Felix Guattari, *On Machines*

Here, me, now

Subjective experience is *phenomenal consciousness* (Block, 2002) and since the standard scientific method relies on an objective account of the mind based on empirical evidence, it can't directly explain it (Chalmers, 1995). Philosophy, instead, has developed different methods to look at the *phenomena* (the things that appear to us) in themselves[19]. At the end of the 19th century, Edmund Husserl's *phenomenology* inquired about the nature of mental content acknowledging the possibility to infer objective knowledge about it and the external world. During the first half of the 20th century, analytic philosophers theorized the *sense-data*, later *qualia*: minimal mind-dependent unities which combined together constitute the whole phenomenal consciousness (Metzinger, 2009). These approaches and the description of the mind portrayed by the aforementioned cognitive revolution involve the *mental representation*[20] of the external world (*representational realism*) instead of direct contact with it (*naive realism*) (Metzinger, 2009). Our perception, is deconstructed, processed in different areas of the brain, and recompose in the world as we experience it.[21]

The contents of our phenomenal consciousness accessible through introspection are resumed in the experience of having a first-person perspective (*me*) on the world (*here*) in a specific moment in time (*now*). Generally, our point of view takes place from within our body, which is itself represented as part of the world, giving us a sense of embodiement, ownership and selfhood, as well as location, presence, and agency (Metzinger, 2009). On the one hand, the 'me' or *self*, as experienced by humans and few mammals, is built on a higher-level of consciousness allowing us to access memories and being projected in the future, using language and logico-mathematical thinking. Turning the first-person perspective inward, this *extended* or *secondary consciousness*, makes us particularly self-aware beings, able to explore our own mental states and to account and experience 'experience' itself. The lower-level, called *core* or *primary consciousness* is common in humans, a large number of mammals, and marine organisms such as octopi, and consist of a more basic form of self-awareness. On the other hand, the representation of space and time persist in most of the species as a basic-level called by neuroscientist Antonio Damasio, *nonconscious protoself* (Hayles, 2013, 2014, 2017)[22]. I will return later on this argument but it is important to highlight that the absence of a consistent subject capable of *inwardness*, makes us doubt on which extent certain animals are able to experience emotions and feelings as originating from within themselves. However, the impossibility to know what it is like to be another living being leaves this argument open to debate (Nagel, 1974, Chalmers&Friedman, 2020).

Given the hypothesis that the brain is the sufficient cause for consciousness to exist[23], whatever constitutes it, must have a sort of correlation with the physical brain. This is what scientists call the *neural correlate of consciousness* (NCC) (Tononi&Koch, 2015), an extremely complex but *"coherent island emerging from a less coherent flow of neural activity"* that than becomes a more abstract *"information cloud hovering above a neurological substrate"* (Metzinger, 2009). Clinical cases and *limit experiences* that are directly accountable (such as neuropsychiatric syndromes, dreams, meditation, use of drugs and so on) help to map

19 Eastern philosophy, already developed a philosophy of mind before Christ that are still particularly relevant for the recent development in the scientific understanding of consciousness (Hayles, 2017).

20 These mental representations can be understood as functional models produced by the evolutionary process and naturally selected for their survival and adaptive value (Metzinger, 2008).

21 In general, this scientific view is based on empirical data implying that the physical reality described by nuclear and quantum physics exists and our phenomenal experience is projected on top of it.

22 Hayles (2017) refers to the works on neurosciece of Antonio Damasio and the Nobel laureate Gerald Edelmann which respectively defines the two levels of consciousness as *core and extended* consciousness and *primary and secondary* consciousness.

23 This is the essential condition for a reductionist theory of mind (Metzinger, 2008).

which part of the brain is activated when the experience of the ‘here, me, now’ happens in different circumstances (Metzinger, 2005, 2009, Tononi&Koch, 2015). In fact, in spite of an *all-or-nothing* process, consciousness is *graded* and *non-unitary* taking place in different phenomenal worlds. If we manage to link a particular subjective experience with a pattern of chattering neurons, we could get closer to solve the hard problem of consciousness. In particular, the first step to explain subjective experience would be to solve the *one-world problem*: *how different phenomenal facts are merged together (world binding) in a coherent whole*, and defining particular NCCs should lead to finding the *global NCC* and the *minimal NCC* necessary for phenomenal consciousness to take place (Metzinger, 2009).

In his book *Ego Tunnel*, Thomas Metzinger (2009) defines consciousness as “*the appearance of a world*” and the brain is understood as a ‘*world engine*’ capable of creating a wide variety of explorable phenomenal worlds. In particular, he focuses on the phenomenal worlds of dreams and *out-of-body experiences* (OBE) in order to develop a functionalist, reductionist, theory of consciousness. These states of mind, where a complete experience of *disembodiment* can be achieved, have led him to develop a particular definition of self, that instead of being a stable instance, is a process running in our brain when we are conscious and turning off when we fall into a dreamless sleep. Exactly like the experience of the here and now is possible because they exist as internal mental representations, Metzinger's self is identified with the *phenomenal self-model (PSM)* created for better control over the whole organism, and the *phenomenal model of intentionality relation* (PMIR) as the model of its relations to others. Although the internal modeling of the ‘here, me, now’ allows a deeper understanding of phenomenal consciousness as simulated virtual reality, Metzinger's claims that "*no such things as selves exist in the world"*. This provocative claim, however, might be misleading in understanding the nature of the ego, which, notwithstanding the perspective of an internal self-model, seems more ontologically rooted when we consider the tangibility of experience itself (Hayles, 2013, Chalmers & Friedman, 2020).

Metzinger, like other researchers, tries to explain why it really looks like we are living in a simulation created by our own brains. While conscious experience seems to take place far away from the physical world, as an indirect representation, it seems to dwell in a place other than the physical brain which instead is the object of study of most of the scientific community. Drawing a liminal space existing between our brain and the physical world and claiming for a reality of the phenomenal world closer to dreams, Antti Revonsuo calls the experience of being ‘here, me, now’ reasonably an “*out-of-brain experience” (as cited in Metzinger, 2009)*.

Engines and experiences

If the actualization of the *computer metaphor*[24] in a computationalist perspective, has its limitations in practice, kept as a metaphor it helps us think about many aspects of our beings. In particular, the difference between hardware and software reflects our struggle to interpret the relation between our body and our mind, our brain and our consciousness. The first part of this text highlights how computers can produce symbolic and sub-symbolic operations, evolutionary dynamics and embodied knowledge, resulting in external behaviors identical to living beings. However, the available thinking machines cannot be said to be conscious. Most evidently computers lack that active individual instance called self which makes a world to appear. But what about the ‘here’ and the ‘now’ of computers?

24 The *computer metaphor*, which compares the brain to a computer, here is emphasized in its ‘metaphor’ instead of its ‘computer’ which instead, when compared to computaionalists approaches, should be intended as ‘computational system’ (Rescorla, 2020).

In her book *Hamlet on the holodeck*, literary criticist Janet H. Murray (1997) develops a theory of new media based on their literary nature. She reports a quote from Italo Calvino's *If on a winter's night a traveler* describing the experience of a writer in front of his typewriter:

> *"Every time I sit down here I read, 'It was a dark and stormy night...' and the impersonality of that incipit seems to open the passage from one world to the other, from the time and space of here and now to the time and space of the written word; I feel the thrill of a beginning that can be followed by multiple developments, inexhaustibly." (as cited in Murray 1997)*

Murray, explains how the overwhelming capacity of the analog text to project the reader in its world, is reconfigured and augmented in new media. Not only can the text be translated into a digital file, displayed and multiplied, but the whole nature of computers' software, where the digital text takes shape, is itself textual. Both the stack of layers of programming language and the binary code dwelling at its foundation are texts expressing meaning. These computers' backstage has been used critically in literature (Hayles, 2004, Goldsmith, 2011), and software art (Cramer&Gabriel 2001, Cramer, 2002) to unveil the textual nature and the conceptual realm of the processes undergoing the *graphical user interface* (GUI) and to compare them to human nature. Referred to as the *Rorschach metaphor* (Nelson 1974, Turkle 1984), the projective character of digital media, is increased by the unique spatial aspects of the software's environment. Often called *cyberspace*, it represents a geographical space where we can move through, in an interactive process of navigation and exploration. Furthermore, the *user/interactor*, active part of this process, triggers certain events to happen in a temporal immediacy:

> *"You are not just reading about an event that occurred in the past; the event is happening now, and, unlike the action on the stage of a theater, it is happening to you.."* (Murray, 1997)

Integrating space and time, software enables a world to be experienced. As the brain described by Metzinger, computers' hardware works as a world engine. However, because of the absence of an internal experiencing consciousness making the world appear, their *'here'* and *'now'* is actualized only through the subjective experience of an external 'me'. Similarly to this view of software as potential worlds, the computer scientist pioneer in educational software Seymour Papert, has developed the concept of *microworld*:

> *"[The microworld is] a little world, a little slice of reality. It's strictly limited, completely defined [...]. But it is rich. [...] The microworld is created and designed as a safe place for exploring. You will try all the sort of things. You will never get into trouble. You will never feel 'stupid'."* (Papert, 1987)

The microworld works as an educational tool helping children to learn how to operate and design multiple contained digital environments. In the long term, this knowledge of different small worlds can be used to create something larger: a *macroworld* (Papert, 1987).

What we call software, is a stack of abstractions relying on each other but, in the end, it is nothing more than electrical impulses happening on the physical level of the hardware, in fact, '*there is no software'* (Kittler, 1992). The same happens for our consciousness, which scientists are continuously trying to reduce to the brain itself, and, to paraphrase Metzinger, *'there is no self'*. However, the influence of software in our society is widespread. The worlds created by software shape the physical world, and, for many aspects, it is increasingly considered a cultural object worthy of being studied in-depth. Something similar is happening to the self which is actually experienced as more than an abstract model switching on and off. It seems to contain the instruments enabling us to transform a meaningless physical world into a meaningful phenomenal universe (Chalmers & Friedman, 2020) worth to be explored and giving us the means to create our complex society. When the *self* interacts with the *self-less* computer, the projective mechanisms of the textual software activate transporting the individual to experience a new phenomenal world. From this view, if the hardware represents the physical level, the software is not a property of the hardware, but represents the possibility of a phenomenal world which is actualized only when experienced by a self. This phenomenal dimension that the software acquires can be described as an *out-of-hardware experience*, exactly because experienced by a conscious subject located outside of the hardware.

However, if the software is experienced *out-of-hardware*, and consciousness is experienced *out-of-brain*, where is subjective experience exactly located? The identification of subjectivity within the hardware is typical of that researcher whose scientific approach negates and reduces subjective experience to a brain's mechanism. They easily tend to alienate their own selves idealizing computers as living organisms and predicting their ability to generate consciousness autonomously. Instead, when the problem is posed in these terms, the individual can claim back its power over the machine in shaping the center of phenomenal consciousness. In fact, the *one-world problem* of subjective experience mentioned before, includes that one world is first needed for consciousness to take place. A first mental simulation is necessary, and then, from this one world, other simulations similar to the microworlds described by Papert can be performed predicting the results of an action or recalling a past event[25]. However, the hardware and the brain are two different kinds of *world engine*. They are two different systems and, even though producing the same results, they differ precisely in substance, structure, and processes (Dreyfus, 1972). When we experience software, a phenomenal world, other than the simulation of our main world, opens in front of us. From the inside of the first world *out-of-brain*, a second world *out-of-hardware* can appear. In computer science, when a system runs a simulation of another system, this is called *emulation*. Given this notion, the brain and the hardware can be understood respectively as a *world simulator* and a *world emulator*, when seen from the perspective of subjective experience.

Extending cognition

To better understand the relationship connecting humans and machines, it is necessary to understand how the phenomenal *'here, me, now'*, the compound of consciousness, differs from the *'here'* and *'now'* of the selfless world of software. This debate has progressed in many directions, however, the foundational elements to understand this relationship have been there all along.

A first connection is already contained in *Erewhon* (1872) the main novel of the aforementioned forerunner of cybernetics, Samuel Butler and in its influence in the work of Deleuze and Guattari. Meant to be read backward, an anagram of nowhere, *Erewhon* contains *the book of the machine* where consciousness was for the first time binding humans and machines. Previously, Deleuze's critique of representation (1968), articulated by his concepts of difference and repetition, will reframe this term not just as a *no-where* but as a *now-here*[26]. Later, in their collaborative work *Anti-Oedipus* (Deleuze & Guattari 1972) they will relate the same term to their concept of *desiring-machine* and Butler's understanding of machines to the *body without organs*. Finally, Guattari (1995) will describe the machine as a *proto-singularity* differing from biological organisms but closely related to their nature.

25 This nested hierarchical structure is common in conscious mental simulations as well as in software, where at the top-level runs the operating system.

26 *"Butler's Erewhon seems to us not only a disguised no-where but a rearranged now-here"* (Deleuze, 1968).

The term, *proto-singularity* suggests a direct link to the aforesaid *proto-self* defined 10 years later by Damasio (2000) as the ensemble of brain devices that continuously and nonconciously maintain the body state within the narrow range and relative stability required for survival, and representing the deep roots of the elusive sense of self' of conscious experience. Still referring to two different domains, technical and biological, the theoretical correspondence of these terms, can be traced back to the offspring of cybernetics of the late 60s[27], and in particular to the researches of biologists Humberto Maturana and Francisco Varela (Guattari, 1995, Hayles, 1999, 2017). They first developed the idea that cognition emerges in living systems from their ability to self-organize as self-contained systems (*autopoiesis*) to then enlarge this position comprehending the sensorimotor capacity of the organism to match and interact with its specific environment (*enaction*). Proposing an alternative to computationalism and connectionism, the enactivist paradigm extends cognition beyond the brain and consciousness into the nonconscious inner processes happening in the organism as a body (*embodied cognition*) that interact with an external environment (*situated cognition*) (Hayles, 2014, 2017, Pasquinelli 2014, Rescorla, 2020). This radical view of cognition can be extended furthermore outside of the body to create frameworks comprehending not only animals and plants but also the technical system, and eventually, natural processes (*distributed cognition*) (Hutchins, 2000, Hayles, 2004, 2014, Pasquinelli, 2014), getting closer to the panpsychist view where the mind becomes a fundamental element of the whole reality.

In her recent work*s* Katherine N. Hayles (2014, 2017), reframes Damasio's protoself as *nonconsious cognition* emphasizing the extension of cognition outside consciousness into embodied and situated processes, and the relevance of the nonconscious as a new cognitive sphere comprehending both biological and technical systems. Furthermore, because cognition presupposes interpretation and production of meaning[28], the nonconscious provides a framework, that she calls *cognitive assemblages*, to extend the social theory beyond anthropocentrism and consciousness, into a cognitive ecology of human and nonhuman *cognizers*. Differing from the unconscious for its inaccessibility by conscious states, the nonconscious posits itself in-between material processes and consciousness, providing the first layer of meaningful representations needed by consciousness to take place. Furthermore, according to new empirical proofs, the nonconscious works faster and can process a bigger amount of information than consciousness, avoiding the latter to be overwhelmed. However, with its ability to choose, at its simplest level between a zero and a one, and perform faster than consciousness, the technical nonconscious can condition our decisions and behaviors making new techniques of surveillance and control possible[29]. From these perspectives, the study of computational media becomes a necessity to complete a coherent map of social interactions and to openly accept their active role in the production of culture.

The framework of the nonconscious cognition developed by Hayles, provides a working model to understand the actual relationship between consciousness and software. In fact, given an extended cognition beside consciousness, on the “biological" hand, we find conscious processes relying on internal, dynamical representations provided by the biological nonconscious. These representations, that are maps of the environment and the body continuously updated in a window of time, provide consciousness with the building blocks of an embodied sense of self and a point of view through which it experiences a coherent phenomenal world. On the "artificial" hand, there is no consciousness and self to reinterpret the representations provided by the technical nonconscious. Furthermore, far from being embodied and situated as biological organisms, the technical nonconscious is an embedded system burnt in silicon, compiling and interpreting lines of text internally stored and manifesting its represented content through an interface. This technical cognitive process

27 This late and last offspring of cybernetics is called *Second-order cybernetics (Johnston, 2008, Hayles, 1999).*
28 The field studying the production of meaning in the biological realm is called *Biosemiotics* (Hayles&Sampson, 2018).
29 This problematic is known as *the missing half-second* (Hayles, 2014, 2017).

happens in *real-time,* like the biological one, and represents the abstract spatial dimension described by its code providing the *'here'* and *'now'* necessary for a world to appear. The software stands for the possibility of the epresentational processes of the technical nonconscious to be extrapolated, internalized and re-represented by a consciousness that integrates its 'self' to experience a new phenomenal world as an *out-of-hardware experience*.

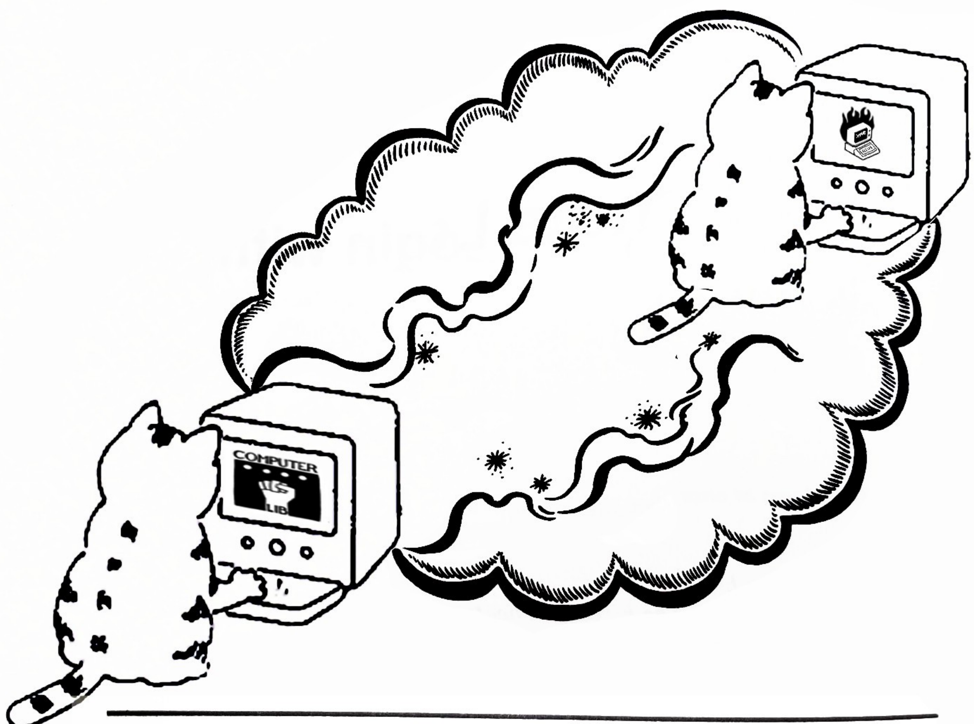

Figure 7.1 Remote login is a lot like astral projection.

Conclusion: a walk through the language maze

The attempts of the proponents of machinic life to build autonomous machines, and the articulations of human-machine symbiosis are essential steps in exploring the processes of cognition. However, these frameworks rely on premature assumptions perpetuated as pretended actualities while they fail to consider the consequences of their claims and products on the broad public. The focus of these disciplines should change because the symbiosis is clearly already happening and it necessarily changes our lives and our societies through a "*control without control*". Indeed, the mimesis of consciousness in technical systems and its underlying faith in a true artificial consciousness must rely on the understanding of the biological consciousness and, eventually, it must be re-framed in accordance with it. Instead of the rush to increase the capabilities of technical systems, developing a science of consciousness is a necessary step that we must take first, and that will allow us to disclose the nature of subjective experience and to re-organize our understanding of the physical world, the biological and technical systems, and the mind.

In the same trajectory, understanding consciousness provides new means to look beyond consciousness itself. It permits us to find the natural position of an elusive object of inquiry which, because it is observable only inside ourselves as subjects, has been used for ages to perpetrate an unnatural anthropocentrism now felt threatened by our own technologies. The extension of cognition outside consciousness, which is already envisioned in Intelligence Augmentation and the cyborg theory, allows us to think of a natural social ecology where different forms of cognition, conscious and not, shape each other in a communal influence. Instead of being a threat, it opens new physical and intellectual relationships with new forms of cognition in and beyond the biological realm.

The interaction between human beings and the technical system through software, as discussed in this thesis, remarks on the validity of such developments and insists on the necessity to continue in this direction. It envisions new ways to articulate the study of software, on the one hand, by standing firmly on the materiality of the physical processes constituting it and reducing it completely to the hardware, and on the other hand, by highlighting how the interaction with a conscious subject permits to rethink software in terms of an experiential world abstracted by the underlying material processes. Drastically different from other phenomena, which fails to provide the complexity of an experiential world, software can be arguably said to augment consciousness instead of just augmenting cognition and intelligence. The technical questions on the validity and consequences of this thesis rely on the next developments in our understanding of consciousness, of the physical world, and of their relationship which still is uncertain. These developments, as many people would like to fantasize, might help us to understand that consciousness can be really instantiated in artificial machines able to 'feel' feelings and perceive themselves as embodied in a physical world. They might help us to consider possible to live inside artificially simulated worlds, which currently are still games that can't be mistaken for real worlds. But right now, we must be able to understand why this is not really happening and why we are still deeply different from machines.

Perhaps, the link between the human and the machine consists of the maze created by their intertwining layers of languages.[30] A *"language maze"* made of verbal and non-verbal languages, natural languages and formal languages, computer's code and machine languages. A Deadalus' labyrinth of material, informational, algorithmic and literary explorable spaces developing in the horizontal and the vertical direction, from microscopic to macroscopic territories. A Penelope's web made of rooms hiding recursive simulations and emulations of other rooms, other mazes and itself. Perhaps, what distinguishes the human from the machine, is the capacity, illusory or real, of being whole with the "language maze" as an infinite space where to build new worlds from scratch.

"The Labyrinth is presented, then, as a human creation, a creation of the artist and of the inventor, of the man of knowledge, of the Apollonian individual, yet in the service of Dionysus the animal-god."

— Giorgio Colli, *La nascita della filosofia (The birth of philosophy)*

30 Language here is intended in the broad sense, with non-verbal language and sensorial perception, to suggest -all the possible signifiers.

References

Bibliography

Asimov, I. (1956) ‘The Last Question’ in *Science Fiction Quarterly*, November.

Alpaydin, E. (2016) *Machine Learning; the New AI*, Mit Press, Essential Knowledge Series.

Ashby, W.R. (1956) *An Introduction to Cybernetics*, London: Chapman & Hall.

Block, N. (2002) ‘Some Concepts of Consciousness’ in *D. Chalmers (ed.), Philosophy of Mind: Classical and Contemporary Reading*, pp. 206-219.

Barbrook, R. and Cameron, A. (1995) ‘The Californian Ideology’ in *Mute*, Vol. 1, No. 3.

Cariani, P. (2010) ‘On the Importance of Being Emergent’ in *Constructivist Foundations* 5 (2).

Colli, G. (1975) *La Nascita Della Filosofia*, Piccola Biblioteca Adelphi, 29.

Chalmers, D. (1995) ‘Facing Up to the Problem of Consciousness’ in *Journal of Consciousness Studies 2.*

Cramer, F. and Gabriel, U. (2001) ‘Software Art’ in Andreas Broeckmann and Susanne Jaschko (eds.), *DIY Media: Kunst und digitale Medien: Software, Partizipation, Distribution. Transmediale. 01*, Berlin: Berliner Kulturveranstaltungs, pp 29-33.

Cramer, F. (2002) *Concepts, Notations, Software, Art.*

Cramer, F. (2002) *Contextualizing Software Art.*

DeLanda, M. (1997) ‘The Machinic Phylum’ in *V2_, Technomorphica.*

Deleuze, G. (1968) *Difference and repetition*, Continuum, trans. Paul Patton.

Deleuze, G. and Guattari, F. (1972) *Anti-Oedipus*, University of Minnesota Press, trans. Robert Hurley, Mark Seem, and Helen R. Lane.

Dreyfus, H. (1965) *Alchemy and Artificial intelligence*, RAND Corporation.

Dreyfus, H. (1972) *What Computers Can't Do*, Harper & Raw.

Goldsmith, K. (2011) *Uncreative Writing: Managing Language in the Digital Age*, New York: Columbia University Press.

Guattari, F. (1995) ‘On Machines’ in *Andrew Benjamin (ed.), JPVA, Complexity*, No. 6.

Hayles, N.K. and Simpson T.D. (2018) ‘Unthought Meets the Assemblage Brain’, in *Capacious: Journal for Emerging Affect Inquiry* 1 (2).

Hayles, N.K. (2017) *Unthought: the Power of the Cognitive Nonconscious*, The University of Chicago Press.

Hayles, N.K. (2014) 'Cognition Everywhere: the Rise of the Cognitive Nonconscious and the Costs of Consciousness' in *New Literary History* 45 (2): 199– 220.

Hayles, N.K. (2004) 'Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis', in *Poetics Today* 25.1, 67-90.

Hayles, N.K. (1999) *How We Became Posthuman*, The University of Chicago Press.

Haraway, D. (1985) 'A Cyborg Manifesto', in *Socialist Review.*

Hutchins, E. (2000) 'Distributed Cognition' in *IESBS.*

Jackson, F. (1986) 'What Mary Didn't Know' in *The Journal of Philosophy* 83 (5).

Johnston, J. (2008) *The Allure of Machinic Life*, MIT Press.

Kittler, F. (1992) 'There Is No Software' in *John Johnston (ed.), Literature, Media, Information Systems: Essays.*

Levine, J. (2009) 'The Explanatory Gap' in *Harold Pashler (ed.), Encyclopedia of the Mind, SAGE Publications*.

Metzinger, T. (2009): *The Ego Tunnel: The Science of the Mind and the Myth of the Self*, New York: Basic Books.

Miller, G.A. (2003) 'The Cognitive Revolution: a Historical Perspective', in *Cognitive science* 7 (3).

Murray, J.H. (1997) *Hamlet on the Holodeck: the Future of Narrative in Cyberspace*, New York: Free Press.

Nagel, T. (1974) 'What Is it Like to Be a Bat?' in *Philosophical Review* 83, October.

Nelson, T.H. (1974) *Computer Lib: You Can and Must Understand Computers Now / Dream Machines: New Freedoms through Computer Screens—A Minority Report*, Hugo's Book Service.

Papert, S. (1987) 'Microworlds: Transforming Education', in *Artificial intelligence and education*, 1, 79-94.

Pasquinelli, M. (2014) 'Augmented Intelligence', in *Critical Keywords for the Digital Humanitie.*

Plato, *Euthyphro*, trans. Benjamin Jowett

Rescorla, M. (2020) 'The Computational Theory of Mind', in *Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy.*

Rushton, S. (2019) *Annotations on Butler and Bateson.Purpose in Animal and Machine*, unpublished.

Searle, J.R. (1980) 'Minds, Brains, and Programs', in *Behavioral and Brain Science* 3 (3).

Tononi, G. and Koch, C. (2015) ‘Consiousness: Here, There and Everywhere?’ in *Phil. Trans. R. Soc.* B 370: 20140167.

Turing, A. (1950) ‘Computing Intelligence and Machinery’ in *Mind 59.*

Turkle, S. (1984) *The Second Self: Computers and the Human Spirit,* New York: Simon & Schuster.

Varela, F.J. (1996) ‘Neurophenomenology: a Methodological Remedy for the Hard Problem’ in *Journal of Consciousness Studies* 3.

Vinge, V. (1993) ‘The Coming Technological Singularity: How to Survive in the Post-Human Era’, in *Whole Earth Review,* Winter.

Wallace, R. A. (1858) ‘On The Tendency of Varieties to Depart Indefinitely from the Original Type’

Weisberg, J. (2012) ‘The Hard Problem of Consciousness’ in J. *Feiser & B. Dowden (eds.), Internet Encyclopedia of Philosophy.*

Other resources

Chalmers, D. and Friedman, L. (2018) *David Chalmers: The Hard Problem of Consciousness*, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=LW59lMvxmY4.

Friedman, L. (2018) *MIT AGI: Artificial General Intelligence*, [video lecture], Available at: https://www.youtube.com/watch?v=-GV_A9Js2nM.

Hayles, N.K. (2018) *The Cost of Consciousness and the Cognitive Nonconscious*, [video lecture], Available at: https://www.youtube.com/watch?v=7iDL9yDH4ko.

Kahneman, G. and Friedman, L. (2020) ‘Daniel Kahneman: Thinking Fast and Slow, Deep Learning, and AI’, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=UwwBG-MbniY.

Marcus, G. and Friedman, L. (2019) ‘Gary Marcus: Toward a Hybrid of Deep Learning and Symbolic AI’, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=vNOTDn3D_RI.

McCorduck P. and Friedman, L. (2019) ‘Pamela McCorduck: Machines Who Think and the Early Days of AI’, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=i6rnzk8VU24.

Musk, E. and Friedman, L. (2019) ‘Pamela McCorduck: Machines Who Think and the Early Days of AI’, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=dEv99vxKjVI.

Musk, E. and Friedman, L. (2019) ‘Elon Musk: Neuralink, AI, Autopilot, and the Pale Blue Dot’, [podcast], *Artificial Intelligence (AI) Podcast*. Available at: https://www.youtube.com/watch?v=smK9dgdTl40.

Metzinger, T. (2018) *Being No One: Consciousness, the Phenomenal Self and the First-Person Perspective*, [video lecture], Available at: https://www.youtube.com/watch?v=mthDxnFXs9k.