

From the Golem to Artificial Intelligence: science in the theatre for an existential reflection

Silvana Barbacci

Master's Degree in Science Communication, ISAS, Trieste

This work analyses how the theme of the creation of thinking machines by man, particularly through artificial intelligence, is dealt with on stage, with reference to three plays addressing different topics and characterised by different types of performance.

This analysis reveals the particular effectiveness of plays dealing with scientific topics, when the relationship between theatre and science results in reflections transcending the boundaries of its contents to address man and his essence and gives voice to the ancient question of the sense of the world.

Introduction

Among the various interpretations of what may generically be referred to as “scientific theatre”, [2] there is that type of theatre which draws from science – or from the foreshadowing of what science will enable man to do – elements for an existential reflection, touching man's most intimate questions: the sense of his being in the world, the responsibility for his deeds.

From this point of view, science does not offer the theatre mere contents, but rather live material generating questions on the sense of history,

of life and death. This is the perspective underlying the analysis of the plays which are taken into consideration in this work: *R.U.R, Rossum's Universal Robots* by Karel Čapek, *Il cervello nudo*, by Giuseppe O. Longo and *I Cinque di Cambridge*, by John L. Casti, in its adaptation for the theatre. These works, each in its own particular way, present the audience with a reflection on the theme of “thinking machines”, the result of man’s demiurgic dream to imitate the act of divine creation. This legend – a constant element in Western history stemming from myth and golemic tradition – has seen a concrete possibility of realisation only with the advent of artificial intelligence. It entails an ancient and archetypical dream which gained new strength every time man imagined through literature, or tried with alchemy and technology, to create beings resembling himself: from Paracelsus’s Homunculus to the literary creatures of Faustus and Frankenstein, from the dolls in Hoffmann’s tales to “A Space Odyssey” and “Blade Runner”, from ancient automata and the talking heads of the Middle Ages, to the perfect creations of Vaucanson and Jaquet-Droz, to the androids and andreids which have inhabited human imagination since the dawn of time. However, less than a half a century has passed since technology started paving the way for men to leave the dimension of imagination and dreams, to overcome what may be considered as “unsuccessful attempts”. Cybernetics and information technology are the instruments which have enabled man to explore and imitate his mind’s world, and to recreate what – from a Cartesian perspective – is considered as the “lofty” part of man: his intelligence and cognitive functions. And while a world populated by artificial beings was becoming more than a mere abstract thought, “thinking” machines were starting to tread the boards of a new theatre: that of a forthcoming future

Artificial Intelligence

AI (Artificial Intelligence) was officially born in 1956 at Dartmouth College, Hannover, New Hampshire, during a seminar: the *Summer Research Project on Artificial Intelligence*, organised by mathematician John

McCarthy and financed by the Rockefeller Foundation. The seminar – as McCarthy himself stated – was based on the conviction that “any aspect of learning and any other characteristic of intelligence may – in principle – be described so precisely as to be simulated through a machine. The discipline of artificial intelligence was “born orphan but with many godmothers: psychology, philosophy, biology, physics, mathematics and engineering”[3]. It came to light in an atmosphere full of expectations and uncertainties at the same time, as the mechanisms of thought, conscience and intelligence were still far from understood. AI therefore appears to be an ambiguous discipline, still overwhelmed by the disquieting shadow of golemic tradition but deeply rooted in the “rigorous” and rational thought of Descartes and Leibniz. It is a branch of knowledge which will lead man to great conquests but also to great failures and disappointments. If, unquestionably, electronics and information technology underlie practical realisation, artificial intelligence also inspired reflections of a philosophical and psychological nature on the mechanisms of the mind and intelligence and on what may be defined as an intelligent behaviour. The first formulations of functionalistic artificial intelligence assumed that the mind was resolvable through a complex series of instructions, of algorithmic functions, that it could take an abstract form, perfectly autarchic and independent of the sentient body. In the past few years, also following a series of clear failures (particularly in the reproduction of language), the importance of the body as a primary perceptive instrument has been rediscovered, and technological development has shifted towards the creation of machines – such as neural networks – taking into consideration hardware, i.e. the physical media, as well as software.

Development of AI in the first two decades was characterised by the “repression” of the body, which is possibly a strong legacy of our culture, starting from Plato or at least from Descartes. According to Longo [10]: “our civilisation developed in the light of the twofold repression of body and genius. Distrust of the body, which has spread through our cultural tradition for almost two millennia, is mingled with the feeling of uneasiness and anxiety procured by genius through its unjustifiable and outrageous unconventionality. Body and genius are the symbols and the moving forces

of disorder, hindering man's continuous effort to establish order and rules in the excessive and chaotic world he lives. With the help of reason and rationality, man is always trying to transform reality, replacing the world – too luxuriant and exuberant – with a more aseptic and less violent one, a measured and controllable world in which he can survive”.

Now, science, and particularly the pure sciences, are based upon the splitting of the *res cogitans* and the *res extensa*, to the detriment of the latter. Starting with Plato, through to Descartes, Leibniz and Kant, man's cognitive activity has been reduced to mere aseptic and acontextual rationality, refusing or considering any sensitive experience as secondary – if not misleading – in cognitive processes. In this perspective, symbols, principles, theories and algorithms are believed to be sufficient for the reconstruction of human intelligence. Again, quoting Longo: “The dream to mechanise thought, thus exorcising the genius' mysterious and suspect powers, has been a constant element in the development of the modern age; indeed, the theatres of memory, the *Ars Magna*, the *Characteristica Universalis*, the Analytic Machine, etc., are but cognitive expressions – more or less refined – providing us in one go with all “true” propositions, with all “exact” results and “demonstrable” theorems mysteriously achieved through the genius. Even Descartes' analytical geometry is a mental prosthesis which – through a series of mechanical devices –permits mediocre minds to demonstrate the most difficult propositions of that discipline, which would otherwise require imagination, intuition and talent”.

The reproduction of intelligence through mechanical or electric systems was based upon the idea that brain functions could be fragmented, atomised, and that thought could be recreated through a series of symbols and rules organised in algorithmic structures. However, this “cold” and abstract intelligence cannot reproduce a great portion of reality, namely anything relating to emotions, feelings, the unconscious, dreams, our historical-biological origins, our living incarnated in a body, communication through language, art and all that refers to the transcendent. This variety of intermingled aspects is not only connected to the non-scientific world: even the science of the twentieth century had to face issues such as complexity, the relation between subject and object in quantum mechanics, the principle

of indetermination and incompleteness theorems, all aspects leading man – still anchored to his claim to give a rational explanation of reality – to a dramatic crisis and a deep feeling of inadequacy. Man’s intellective and abstract tools, though complex and articulate, are therefore frail and need to combine with “ancient traditional components”, having to do with the earth, the body and the physical dimension in general. These elements provide an immediate, direct understanding of the surrounding world, acting on instinct, intuition and the senses. The kind of knowledge that passes through these “channels” is more archaic and deep rooted than rational knowledge, which is more recent and frail, depending on the other types of knowledge but with a less immediate effect, having undergone a process of abstraction. “The first type of knowledge – states Longo [7] - corresponds to ancient mental maps, which evolution has brought to a deeper level and rooted into the individual’s biology. On the contrary, maps relating to the second type are superficial, must be recalled with a conscious effort or created *ad hoc* in case of necessity, and only exist at a rational level. The longer the evolutionary history of the map, the deeper its place inside man and the more unconscious and immediate its use. Maps of the first type might be defined as "natural", while those of the second type as "cultural". Again, Longo writes that: “the history of western culture is basically a constant effort towards transferring or *translating* the various types of knowledge [...] from biological knowledge incarnated in the body (which is, in its turn, immersed in the environment) into disembodied rationality. This effort is reflected by the functionalistic or physical-symbolical structure of artificial intelligence [...]. The question which arises is: to what extent is this transfer possible? Though eager and strenuous, this effort is hindered by incompleteness: the obstacle to all translation processes. There always remains a trace, an incurable scar reminding us of the impossibility of translation, in that it would like to be or should be a total superimposition of the world on itself”.

Theatre and imitational creation: analysis of three plays

The theme of second rate creation, or imitational creation, connected with the theme of artificial intelligence – namely the construction of “intelligent” machines, capable of thinking and possibly feeling emotions as men do – has been a frequent element in scientific literature, science fiction and theatre, the latter being the privileged place for staging and reflecting upon the existential themes which are closely related to the creation of thinking and sentient machines. Within the territory of imitational creation, as Giuseppe O. Longo says [14], “the two opposite aspects of manifest diversity, arousing astonishment, horror and alarming resemblance, are a source of ambiguity and complex ethical problems, confronting us with the responsibility of the creator himself: indeed, the enormous complexity of the creature thus created, deriving from its perfect resemblance to its model, arouses questions concerning its possible feelings and reactions. The psychology and sociology of automata, androids and cyborgs are some of the most interesting issues of modern science fiction and possibly one of the most complex problems of the near future. Why give life to creatures which are so similar to human beings that they are capable of suffering? Their sufferings, often stemming from the awareness of not being completely equal to men would be a sad after-effect of our power of creation”.

The following portion of the text will focus on the analysis of three plays which are centred on the problems related to machines imitating human faculties, to imitational creation and artificial intelligence *tout court*. More precisely, the plays are *R.U.R. Rossum's Universal Robots* (1920), by the Bohemian writer and playwright Karel Čapek who, in the first half of the twentieth century, imagined what might happen in a society in which mass-produced robots overwhelmed men. Secondly, *Il cervello nudo* (1998) by Giuseppe O. Longo, in which the author dwells upon the inner drama of a scientist who has created “machines intelligent enough to understand they are not intelligent enough”, and has therefore given life to “indefinite” and suffering beings. Finally, the theatre adaptation of “I cinque di Cambridge” by John L. Casti, edited by Luca Scarlini, in which the author wonders about the effective possibility of creating an intelligent machine. In an imagined

meeting, Turing, Wittgenstein, Haldane, Schroedinger and Snow discuss whether a machine will ever be able to think and have a conscience.

Whilst the first two plays deal with the existential issues deriving from man's demiurgic act - through a playwright's science-fictional prefigurations in *R.U.R.*, and through a scientific drama urging for a reflection on the implications of a certain use of knowledge in "Il cervello nudo" - the third work is original both in the type of performance and in the way the themes are dealt with. Indeed, "I cinque di Cambridge" focuses on philosophical dialogue, and the play - concerning the possibility of creating intelligent machines - sets aside the dramaturgical plot based upon narration and action and mingles with the tradition of show-conference, particularly widespread in England.

These are not the only examples of how the issues concerning imitational creation and thinking machines were brought to theatre, but their particular significance lies in the fact that, through a different staging, different contents and perspectives, they show the twofold relationship of science and theatre [2]: theatre watching science and science watching itself on stage.

***R.U.R.* Rossum's Universal Robots**

The play *R.U.R.* Rossum's Universal Robots, by the Bohemian writer and playwright Karel Čapek, dates back to 1920 and is also known for its introduction into everyday language of the word "robot", a Czech word meaning "android", "artificial worker", and which Čapek derived from the Slavonic root *rabotat* – to work. The word *robot*, with its current meaning, was absorbed by western languages after the London success of the play in 1923. The plot deals with the extraordinary discovery of scientist-philosopher Rossum (*rozum* means "reason"), a character who seems to come straight out of that abundant crowd of "wise madmen", a typical theme of the central European expressionism of the time. Rossum finds the formula of a chemical substance that brings matter to life, and his nephew, an engineer, decides to use this discovery to start mass production of automata.

Thus, he founds the factory of Rossum's Universal Robots, artificial beings created to cancel work from men's lives. At the beginning, in the prologue, the factory is managed by a series of characters having evocative names: Harry Domin (Dominus), the general manager, engineer Fabry (Faber), the general technical manager, doctor Gall (Galenus), the physiology and research manager, doctor Hallemeier, the director of the Robot psychology and education institute, consul Busman (Businessman), the sales manager, architect Alquist (Alchemist), the head of constructions. As a consequence of a planning mistake by doctor Gall, Robots become too similar to human beings and when they eventually rebel, they kill the holders of power and of the formula to which they owe their lives. Only architect Alquist survives, and the Robots want him to find the secret which has been burnt by Helena, the factory manager's wife, before dying. The drama ends with the ominous undertones which characterise the whole plot, suggesting a return to nature through the blossoming love between Robot Helena and Robot Primus, which will allow automata to live forever, without any formula.

Robots were conceived by Rossum, the "great philosopher" who retired to an island to study marine fauna and, while trying to imitate "protoplasm", found a substance behaving like live matter, but with a particular chemical composition. In a Faustian impetus, Rossum wanted to imitate the divine creation, to "oust God with science, he was a terrible materialist [...]. All he wanted was to demonstrate that there was no need for God". But Rossum had no intention to exploit the artificial beings he had created. His nephew decided to start mass-production of automata: "It is absurd to create a man in 10 years. If you can't create him more rapidly than nature does, we may as well not bother about all this stuff". The description of these characters, still topical nowadays, underlines the transfer from science – through the negative connotation of challenging nature and even God, and not as love for furthering knowledge – to techno-science for the production of inhuman, dehumanising objects. This, indeed, is Domin's description of young Rossum: "He embodied the new era: the era of production following the era of knowledge". But what was Čapek's idea of Robots? This is how he described them: "Robots are not men. From the mechanical point of view they are more perfect than we are, they have an

extraordinary rational intelligence, but have no soul. [...] The product of an engineer is, technically speaking, more refined than the product of nature". Automata are mechanical beings that "get used to existence". They do not have a will of their own. No passion. No history. No soul. [...] Robots do not feel love for anything, not even for themselves. Do they feel hatred? I do not know: only from time to time... [...]. Sometimes it is as if they broke down. Something like epilepsy. It is called the Robot convulsion. All of a sudden, the Robot breaks everything within its reach, clenches its teeth... and so it must be destroyed. Possibly, the organism simply goes wrong". Though created without any feeling, they can suffer. This is for "industrial" reasons: if robots could feel no pain, they might act dangerously, get damaged and become useless. As concerns the soul, Robots have none: this keeps manufacturing costs low. Automata are created" says Domin "for men not to work anymore. Everything will be done by living machines. Men will only do what they like. They will live to achieve perfection. [...] Maybe terrible things will happen at first [...] and there will be nothing to do about it. But then the slavery of men to other men will cease, and so will the slavery of men to matter. Nobody will ever pay for bread with life and hatred anymore. You will no longer be a worker, you will no longer be a clerk; you will no longer mine coal and you will no longer work someone else's machine. You will not kill your soul with the work you have cursed [...] You will be free and sovereign; you will not have other tasks, other work, other concerns than to follow your own perfection. You will be the master of creation".

The critical conscience of the drama, stressing the destructive nature of the plan to create artificial beings is Nana. Quoting Helena: "The voice of the people speaks through Nana's words, thousands of years speak through Nana's words". It is Nana who, speaking to the Robots' creators, says: "It is the end of the world. For this devilish pride you have dared to create like Our Lord God. It is a sacrilege and blasphemy, you want to be like God. And as God chased man from Eden, He will chase him from the whole world!". In the course of the drama, Helena and Alquist become aware of a critical realisation. "Oh, Madam", says Alquist, "this is an accursed paradise! Helena, there is nothing worst than giving men Heaven on earth! Because women have ceased bearing children! Because the world is Domin's

Sodom!"). When the Robots rebel, Alquist denounces the real dream of the two Rossums: "the old man was thinking of his evil manoeuvres, and the young one was thinking of millions. And it was not even a dream of the R.U.R. shareholders: they dreamed of dividends. And because of their dividends humankind perishes". "Who cares about their dividends?" replies Domin. "Do you really think that I could have worked for that? I did it for myself, do you understand? For my own satisfaction! I wanted men to be masters! I did not want them to live for a piece of bread! I wanted no soul to grow stupid for someone else's machine, I wanted nothing, nothing of this damned social flotsam to remain! Oh, how I loathe humiliation and pain! How I hate poverty! I wanted new generations!". While no comment is made by Gall, Hallemeier, Fabry and Busman on the meaning and consequences of their deeds, Alquist - who to an extent represents \square apek's thought - bursts into a violent accusation: "I accuse science! I accuse technology! We are all guilty! We have killed men for our own megalomania, for someone's profit, for progress, I do not know for what formidable things! Here you are now, having to come to terms with this greatness of yours! No Gengis Khan has ever built himself such a huge pile of bones!". After the Robots' rebellion - during which they kill their own "creators" - the second Robot explains what has really happened: "We were machines, but fear and pain gave us [...] gave us souls".

The Robots really seem to feel a gleam of the soul which finds its expression through Primus and Helena, the Robots that Alquist is supposed to use for his experiments to find the secret of life that got lost during the rebellion. Primus and Helena protect each other to avoid dissection for Alquist's experiments and the latter, seeing their love, decides to let them go, blessing them as though they were the new Adam and Eve. The drama, full of golemic and alchemistic-magical references, ends with a tribute to life and love. Alquist reads a few verses on the creation taken from the Genesis and says: "Rossum, Fabry, Gall, you great inventors, what did you invent that was great as compared to this girl, to this young man, to this first couple, who exalted love, love, love between a man and a woman? Nature, nature, life does not perish! [...] Once again everything starts from love, from being tiny and naked. This beginning is rooted in the wasteland, and everything we

have ever made and built will be useless; cities and factories, our art, our thoughts. Yet love will not perish! Only we will be dead! Houses and machines will be destroyed, systems will break down and the names of the great will fall like leaves, and only you, love, will bloom over the ruins and sow the seed of life in the wind. Now, Lord, forgive your servant, as my eyes have seen, yes, have seen your redemption through love, and life will not perish! It will not perish! It will not perish!”. Alquist – almost Čapek’s *alter ego* – stresses the value of balance and moderation, of the respect for nature. When, after murdering men, the Robots realise they are also doomed to die, the prevention of extinction will not be provided by the intellect, but by love. With regard to this ending, Čapek wrote in a letter to his wife: “I felt sick, Olga, and for this reason in the end I looked frantically for a solution of love and settlement. Do you think it is believable?”. Another aspect of the golemic tradition can be found in the epilogue: the sexual awakening of automata expressing their wish to become human. If, on the one hand, the theme of the Robots’ rebellion echoes the Golem legend - the creature rebelling against its own creator – on the other hand it portrays the rebellion of the oppressed, a subject along the lines of popular dramas of Čapek’s epoch. The drama is both a warning to the new technological society and the expression of the author’s aversion for “the rhetoric of collectivism, for class hatred and for totalitarian ideologies, breaking up the world in the name of a deceptive transformation. If the repulsive spread of salamanders (in “La guerra delle salamandre”, 1936) reflects the spread of the Nazi regime, the Robots’ rebellion reveals indirect references to the Russian revolution. [...] The violent manifestos of Čapek’s androids echo the slogans of the propaganda and the edicts of Bolshevism”, as Angelo Maria Ripellino stresses in the note [5].

Čapek did not trust violent reforms promising bright scenarios for the future and did not cherish many hopes as far as changes were concerned. For this reason, in his text – through Alquist’s voice – an inclination towards balance prevails, together with respect for nature and other human beings. He stresses that love becomes man’s only possibility of redemption.

Il cervello nudo

Il Cervello nudo, written by Giuseppe O. Longo for a “science-theatre” project and staged by the theatre “La Contrada” of Trieste in 1999, is a work which may find its place in the framework of a theatre that draws from science and technological development the reasons for a new reflection on man. The author, a professor of Information Theory at the faculty of Engineering of the University of Trieste, narrator, writer and essayist, is interested in epistemology and in the relation between technology and society. In *Il Cervello Nudo* he exploits theatrical language to reflect on the existential questions raised by the theme of “thinking machines”. Concerning the relation between science and theatre, Longo says: “I was led to deal with science in some fiction and theatrical works of mine by the conviction that science poses grave problems to man, which are of an ethical and existential - and not scientific - nature. Scientific problems are never serious: they can be interesting, banal, curious, stimulating, but they have nothing to do with man in his entirety. They are almost always superficial problems, which become central when they touch on more inner problems: the problem of the constitution of matter becomes a serious problem when a weapon causing suffering, death and destruction is created.” As regards the relation between science, fiction and the theatre, Longo continues: “Science poses problems to man when it becomes bound up with life; and fiction, where life is narrated, can then become the arena for discussion - and therefore criticism - of science, positive and negative, presenting its merits and its limits. It has nothing to do with the explanation of science, or rather: the ways in which science is explained have been laid down over the centuries and have led to a kind of stereotyped summary, dry and bare, which is very far from the richer and more flourishing forms of traditional fiction. But science is fiction, too, even if implicit and almost disenchanting. It simply develops in other ways and pretends to have nothing to do with life and its creator. So, a story can be a place for reflection on science, like the theatre, with the difference that in the theatre the body has an absolute importance, even greater than the word. I do not believe that theatre can be a place where scientific ideas can be communicated: today, scientific communication takes place in stereotyped

forms, far removed from the poetic afflatus; science demands precision and a single voice, while poetry, theatre and fiction live on ambiguity and continuous transgressions of time and space; in theatre, time and space are not homogeneous or linear and isotropic”. Again quoting Longo : “Theatre does not need science, but it needs the body-soul of man. If man is, by chance, a scientist, it is all right anyway, because even a scientist can be the protagonist of a (metascientific) drama. Basically, the link between theatre and science passes through the man-scientist and not through the ideas and concepts of science. A theatre wanting to stage the concepts of science would be very boring and ineffective (the case of cinema is different, since it has other technical resources of animation and so on). That is the reason why in *Il Cervello nudo* I tried to represent the drama of the characters, their tormented existence, their dreams. These dramas and dreams have to do with the scientific adventure”.

The story of *Il Cervello nudo* takes place in an undefined remote area, on a cliff that falls sheer to the sea, where professor Arcularis, a genius of artificial intelligence and the first creator of intelligent machines, had been found a hospital by the government. After creating those machines, the professor has been infected by an unknown disease, a sort of madness for which he is being treated by a strange physician, doctor Krajlevic, who tries to discover the causes of the strange disease through a series of exhaustive analyses and radiographies, thus also insuring himself a brilliant career.

In an undefined period of time, as the author says, “between night and day”, in a dimension that, even if real, is marked by a strong dreamlike component, the story of the characters waiting to meet Arcularis takes place: Marion, the daughter, whom the father has not wanted to see for years, Arne, who was the Arcularis’s personal assistant, and Bonaldo, a journalist waiting to interview Arcularis. They seem to be waiting for a time that will never come, since Arcularis is now in a different space, where only sea breezes, the vague scents of the East and the voice of the radio giving news to mariners exist. Arcularis seems to have found refuge in this almost autistic world of complete isolation to flee from the suffering that is progressively killing his head day after day, tormenting him with memories like a scalpel. Arcularis, like the characters of *R.U.R.*, was guilty of a serious act of *hybris*. He too

wanted to be a creator, and caused sufferings giving life to “machines intelligent enough to understand that they are not intelligent enough”. Now he can free himself from his ghosts, the spectres haunting his mind, only by seeking refuge in a dimension which is abstract from reality, in a world of his own. The drama, drawing inspiration from two tales previously written by the author, *Machina Dolens (Painful Machine)* and *Avvisi ai naviganti (Shipping News)*, highlights some of the questions man has to face when he has the possibility of going beyond his limits, when science provides him with the power to create, which is always connected with the power to destroy. The characters alternating on the stage wonder, in a context of mutual incommunicability, about the meaning of “progress”, about the value of memory compared with the value of oblivion, about narration and cold rationality as opposed to the need for myths coming from the night of time and the dawn of man. These make the audience reflect on the meaning of disease and pain and on the responsibility of action.

The drama sees two characters who, in different ways - one more ingenuously and the other more ideologically - support the value of “progress”, scientific discoveries and technological achievements, all seen in a positivist sense.

Bonaldo is enthusiastic considering the prospect of conquering intergalactic space: “Today man feels lonely, Marion. He is tired, his path is too difficult, he needs strong fellow travellers... hard, tough, tireless,...the machines...no more blood but electromagnetic fields, no more flesh but silicon, no more eyes and nostrils but diodes and integrated circuits... in the future machines will take the place of man and will prolong the mission...”.

Such is the confidence of Bonaldo in technology, in man’s aseptic ability to create a soulless world, the fruit of a history written by men who, counter the feminine ability to procreate (creation of the first order), with a second order, artificial creation, the birth of the mind of machines which are then left, on the threshold of awareness, to their own pain. Marion, who carries - in Jung’s sense - the *feminine* values of the bond with the earth and of the sufferings deriving from denied fertility (Marion aborted the life she had conceived together with Arne), replies to Bonaldo: “But will these machines sing around winter bonfires, will they contemplate sunsets above

the oceans, will they dream of gnomes and unicorns in enchanted forests? ... Will they listen to tales, will they compose the poems of the world?" Bonaldo answers: "Poems and fairy-tales will be of no use... there will be no more mysteries, we will not be the victims of superstition and trickery... man will understand everything and will build a better future...". And Marion: "But that better world will not belong to him, it will no longer be his world, it will be a precise and punctual world where intelligent and purposeless machines will perpetuate empty rituals... the emotional apathy of its poor automata will rule...". The antagonistic dialogue between Bonaldo and Marion underlines the dichotomy between two opposing visions: cold rationality against emotional life, theoretical law against myth, algorithm against rite, the expression of a religious sense connected to the acknowledgement of mystery in the world. The rational systems created by man cannot provide an answer or an explanation for the existence of affections, of feelings, of sensations, of the unconscious, of the channels of communication passing through the paths of intuition and instinct. They do not take into consideration "that obscure and blinking place we are continuously attracted to, the place of the first germination, of the dead, of premonitions, of the kindred, of children. A place from which we have worked hard to leave to free ourselves from the human condition, but which does not stop calling us with a voice that can be heard when the heavy clamour of thought and of the machines tones down or stops. Rationality rejects this place, that delicate and sensitive point where we find ourselves and become what we are, and reflects the elusive and singular character of our humanity. In ourselves we bear the mark of everything, of the shadow we came out of, as well. What does the voluntary detachment from our germinal line entail? The formal reconstruction of the world would clearly mean a rejection of our psychobiological history, of the body and its fundamental needs, a decline in the unconscious and a denial of femininity" [10].

Krajlevic is the character who supports the absolute positivity of scientific progress with a strong ideological bent. Krajlevic replies to Arcularis, tormented by the sufferings caused by the disease that is tearing him apart. "Do not be afraid, professor. You are a great man. In your life you

have done a lot for mankind. You have given a fundamental contribution to research on artificial intelligence”. But Arcularis answers: “Yes, that is true. I created those extraordinary machines... just imagine, the first intelligent machines!... But I caused a lot of suffering, too! I can find no peace. You cannot imagine how much my creatures suffered”. Along with his artificial creatures, Arcularis wonders why, his natural “creature” always comes to his mind, his daughter, whom he had not only forgotten when he was doing his research but had even disowned because she had conceived (and aborted) a baby with his assistant Arne.

This is the drama of Arcularis who seems to draw, from his pain and disease, the signs of a more ancient wisdom, of a knowledge “prior” to the birth of computer rationality. Arcularis feels something that comes from his right hemisphere, “the hemisphere of children and gods”, where there seems to be a reference to the bicameral mind mentioned by Julian Jaynes [6], and his memories intertwine with suggestions coming from the voices of the sea, from the news to mariners, from the scents and images of a far East. “Something begins to move inside the heart, a startle, a memory, a tribulation... evil flowers... creatures that were never born ...”.

The dialogue between Arcularis and Krajlevic develops into the dialectical confrontation between the now sceptical scientist and the utilitarian physician. Arcularis: “We are doing everything to be replaced by machines! My colleagues continue indefatigably to follow the path of artificial intelligence, they make androids and andreids who fall in love with each other and punctiliously imitate the acts of love... can’t you see? They caress, kiss, copulate, but it is all a fake, because they cannot feel anything like love... Simulation, they call it... And they make automata that one day will overwhelm us and will leave for other planets, to colonise them and spread all over the universe this desperate yearning to reconstruct the world. What leads us to do this? Somebody once said that civilization is a form of infection... that may well be true...” Krajlevic: “Yet, what cannot be denied is that research on artificial intelligence has borne and will continue to bear important results. There are many things that man is forced to do that tomorrow machines will be able to do. Not only manual work, but also intelligent work...”

Arcularis: “You know, doctor, the human brain is unstable, it is the centre of continuous catastrophes we perceive as thoughts, crazy and sublime ideas... the neurons of a certain area inflame, they twist like big worms and we say “an idea has come to my mind”, and we begin to work immediately to implement that idea, to make a machine that can move mountains and fill oceans, and in this way the imbalance of our brain spreads to the world, and the world is in danger... we divert rivers from Siberia! We take icebergs to the Equator! We send thousands of satellites armed with mirrors into orbit, so that our cities can always be lit by an implacable reverberation!... Madness...[...]. Everything is contaminated, devastated... I can only see twisted pylons, filling stations blackened by fires, abandoned airports, smoky factories, quiet deflagration at the borders of deserts, poisonous smoke coils rising from rain forests... and all this is made by anonymous industrious servants, technical gnomes, hard-working insects re-sinking into the nothingness from which they came out for a moment to bring their contribution to the decay of the world... and this contribution, even if minuscule, even if extremely small, added to all the other tiny contributions, corrodes and damages...”.

Krajlevic: “The organs of our body will be made of resistant plastics and optic fibres, our brain will be integrated with active small plates increasing our intellectual abilities, we will know everything, we will remember everything!”

Arcularis, even though he does not want to repudiate science, believes in moderation. The pain, which provided him with a visionary power, leads him to say: “Standing out from a threatening sky, I see the spiky shape of the first almost intelligent machines, the rejects, dinosaurs abandoned in the overwhelming river of time, creatures full of resentment at being called for a mission whose sense they never understood. Replacing nature, we too began to fill the path of evolution with corpses. To create a superior being: yes, but this does not justify the abortions, the screaming creatures, the discarded flesh thrown back into the melting pot, all the pain, the hatred, the curses inflaming the air”.

After Krajlevic and Bonaldo have yielded to a form of delirium extolling “the wonderful and progressive future” of humanity, as opposed to

Arcularis and Marion who believe that only a return to nature can save man, the play comes to an end with the silent listening to the voice of the sea, which may contain the secret of life.

Gianni Zanarini's [21] comment on the play concerns the analogy between technologic development and the evolution of living beings: "We may wonder: when will there be a machine capable of thinking for itself in the world, capable of self-consciousness and reflective thought? But at this point other questions hound us. What is self-consciousness? What is reflective thought? Are they spiritual qualities whose origin cannot be found in the underlying material bases, as complex sciences seem to suggest, are they entirely new abilities that can emerge all at once in sufficiently complex and plastic systems? In the first case, we find ourselves before a reassuring philosophical barrier between us and the machines; in the second case, we can fantasise that one day machines will ask that we explain why we created them and made them evolve, leaving them on the threshold of humanity, with their silicon flesh and their painful and unachievable aspiration to another life". The questions that thinking machines pose to us, Zanarini continues, "do not only regard our *hybris* of unprepared demiurges seized by a delirium of omnipotence, but rather they touch the very meaning of our human adventure: our ability to understand and give meaning to life".

The Cambridge Five

I Cinque di Cambridge (The Cambridge Five) is the theatrical adaptation of the novel by the science writer John L. Casti edited for Mantua's "Festivalletteratura" by Luca Scarlini in 1999. In his note to the text Casti stated that his work "is not really a novel but a work of fiction, or, rather, a work that belongs to that new genre I like to call 'scientific fiction' ". The appropriate Japanese word in this case would be *shoshetsu*, which is far richer in shades of meaning than "novel". Despite the presence of some fictional elements this genre goes well beyond mere chronicle, for it attempts to transpose in a fictitious scenario the intellectual and cognitive questions debated by those who are busy shaping the scientific knowledge and the

technology of the future. [...] The scientific fiction aims to determine in what way decisions taken in the past contributed to shape our world, and how today's decisions are going to affect our future.

“The Cambridge Five” portrays a fictitious dinner during which the novelist and physicist C.P. Snow, who talks about the split that exists within the Western civilization between science and humane learning, invites Alan Turing, J.B.S. Haldane, Erwin Schroedinger and Ludwig Wittgenstein to discuss whether machines will ever be able to think autonomously. The adaptation devised by Scarlini relies on the English acting tradition within which plays are developed somewhere between literature and theatre, a kind of “show” where non-professional actors read out dialogues. Scarlini described it as “an event in the theatre”, or alternatively as a conference with six leading voices, namely those of the scientists and scientific philosophers (Roberta de Monticelli, Giulio Giorello, Michele di Francesco, Roberto Festa, Pietro Adamo) who willingly accepted to partake in the “reading” while fulfilling the typically British rite of having tea. Scarlini himself acts as master of ceremonies, and music of the time is played in the background.

This one-hour performance is in the form of a “conference-play”, commonly used in Britain to debate issues both scientific and humane. Conceptual and scientific matters are staged without resort to a real dramatisation. Scarlini's starting point was the idea of “science on stage”, and during the staging he took particular interest in the cultural climate of the time and in the characters themselves, who were intriguing not just because of their intellectual activity, but also because of their whims, like Wittgenstein's singular passion for westerns and musical comedies. “I wanted to give the impression that this Cambridge meeting was real” Scarlini says “so I created a “conference-play”, an event, a staged entertainment with a subject and people able to discuss it. Every so often, in fact, I indulge in creating “opportunities” through plays that rely on deep reflection. I was also interested in finding out how some people imagined certain things at a time when Artificial Intelligence was just about to officially become a branch of learning. I was not so much interested in the matter discussed in *The Cambridge Five* as in how some people had imagined the world that was yet to come.” The play is set at *Christ's College* in Cambridge in 1949, and the

most heated discussion within the debate develops between Turing and Wittgenstein and their antithetic theories. The reflection on the nature of language, in particular, is what most markedly differentiates Turing's thought from Wittgenstein's. Turing believes that intelligence can be reproduced symbolically and algorithmically, and is consequently optimistic about the question of whether a machine could possibly think. Wittgenstein, on the other hand, claims that language is a social reality open to everybody with rules that have been codified through human interaction. It is therefore not something abstract that can be decodified and recodified just through sheer reasoning.

His thought is thus summarized by Haldane: "If I understood him well, he reached the conclusion that it is not possible to discuss our knowledge of the human mind in the old Cartesian way. The content of the world can no longer be assumed to be of two completely different kinds: an outer world made of solid bodies that are visible in space and time, and an inner world made of objects and feelings. Indeed, according to the new language theory these two worlds overlap very closely, and it is not possible to discuss thoughts and inner feelings as separate from their visible signs in the various situations." It may be pointed out here that what Haldane refers to as new language theory is in fact "new" only as opposed to the theory enunciated in the *Tractatus Logico-Philosophicus*, according to which the world was composed of simple objects that were interrelated independently of mankind and language. Wittgenstein therefore believes that the various situations, our own behaviour and history are of prime importance in the understanding of how we speak of our mental sphere.

Turing, on the other hand, was at the time elaborating the theoretical and practical basis of a calculating machine, and saw calculation as a mere manipulation of symbols, which is exactly what he wanted his machine to do. The next step, however, led to the hope of being able to create something far better than that: a machine that could think like a human being. The constituent elements of his calculating machine, he explained, are the memory and the executive unit. The executive unit draws on the data held in the memory, follows a sequence of instructions - the program - to do the calculations, and stores the result back into the memory. Turing

consequently drew an analogy with the human brain by claiming that its neurons could be thought of as switches that can either be on or off. The similarity between the processes of memory and the modification of the configuration of the neurons on the one hand, and the very same processes in a calculating machine on the other, led Turing to believe he could actually build a machine that thinks. “The only obstacles appear to be technological, not logical”. He then referred to the results the neurophysiologist Warren McCulloch and the mathematician Walter Pitts published in 1943 that demonstrated that the way neurons linked to other neurons could be imitated by means of purely logical elements. This model was based on the assumption that each neuron was activated and subsequently discharged in favour of another neuron exactly in the same way as a logic sequence can entail the truth or the falseness of any other proposition. The correspondence between neurons and logic can be exemplified by signals that may or may not pass through a circuit. Wittgenstein, however, counted that the complexity of the real world could not be reduced to a law of nature: “The whole modern notion of the world is based on the illusion that the so-called ‘laws of nature’ may explain natural phenomena. [...] . A name for that part of fish known as ‘fillet of sole’ can only be found within the context of a language that has already developed and in which rules exist to differentiate objects, use names and carry out actions. These criteria are not to be found within the logic of machines, tapes and codes, but are part of the daily practice of a linguistic community. A sequence of lifeless symbols cannot become imbued with this kind of meaning simply by creating a set of rules that show how to turn that sequence into a new one.” Indeed, in Wittgenstein’s eyes meaning derives from the social use of the natural language. Despite Schroedinger’s suggestion to endow the machines with sense organs that might enable them to learn, Turing remains persistently in favour of a basic form of intelligence and of the possibility to reproduce a mind without a body, or, in more modern words, a *software without hardware*. This concept was the basis for the “strong” artificial intelligence and lost ground some years later when it had to face up to its inability to reproduce the human language. As the idea of a strong artificial mind based on symbol manipulation and on the assumption that the human thought could

be reproduced algorithmically declined, research on the neuron nets was once again vigorously promoted. These are mechanisms that stimulate the activity of the neurons, are closely interlinked and work in parallel, as opposed to the serial system of the symbolic AI models. Neuron nets also have powerful retroactive mechanisms, rely on the learning process and can thus interact with the outer world and “learn”, whereas the symbolic AI perceives the outer world as interference.

It is not by chance that the artificial intelligence achieved its most brilliant results in the solution of those problems that require a minimal interaction with the outer world, such as playing chess, cryptography and mathematics. It is for instance far easier for a machine to win a game of chess against the greatest ‘human’ champion than acknowledging the fact that a chess-board is such, and it is interesting to notice that when Kasparov was defeated by the computer called Deep Blue in 1997 he remarked that he felt as if he were playing against an intelligent mass.

The five characters debate the issue of the mind-body dualism, a recurrent theme throughout the development and construction of the AI machines both at a philosophical and at a practical level. The Cartesian hypothesis regarding the isolation and inferiority of a *res extensa* as opposed to a *res cogitans* - one of Plato’s theories subsequently adapted to the Christian culture that subordinated the body to the mind greatly contributed in shaping the western culture - has now been demolished through common sense. It is nonetheless interesting to notice how in “2001: A Space Odyssey”, a film shot in 1968 when strong artificial intelligence was all the rage, Hal 9000 says as he is about to ‘die’: “My mind is fading away, I *feel* it”, thereby stating through a kind of oxymoron that the fading away of his mind was not a rationally acquired datum, but rather a physico-emotional feeling.

All that was, however, yet to come at the time of the dinner in Cambridge. The future was being planned while the relevant technology was still lacking: the transistor was invented in 1947, and the first integrated circuits came out in 1958. Turing was, therefore, still right when he claimed that “what really matters is what the brain components, i.e. the neurons, actually do, and how they are interlinked. These are the functional and

structural aspects that give the brains its cognitive power. I am convinced that if we were to create electronic neurons interlinked the way they are in the human brains, we would get an electronic device endowed with exactly the same rules the human brains follows to think and act.”

The five characters agree that in order to match human intelligence a machine ought to have the whole human sensory system to be able to interact with the outer world and to learn just like children do.

The discussion then moves on to define what conscience is, and various issues concerning life, the soul, human thought, the definition of the concept of person and social behaviour, and the influence of culture on behaviour in general, art, and so on, are also brought up on the way. The meeting ends with a summary of the stance each character has taken on the matter of whether machines can think. Wittgenstein proves to be a staunch opponent: “Human thought is closely linked to language, which in turn is a direct consequence of a shared form of life, human life. No machine, no matter how craftily built, will ever be able to share that form of life for the simple reason that it is a machine.” Schrödinger still has not made up his mind. He sees no technological or logical obstacle to the possibility of creating a machine that can think, but he does not see a point in it either. In the end he takes Turing’s side, for he sees no reason why a machine that leads us to believe it thinks like a human being could not be possibly built. Haldane, on the other hand, while having no problem in admitting that a machine able to perform a few conjurer’s tricks could be built, doubts that such a fundamentally human characteristic such as cognitive ability could be transferred into it that easily. Haldane believes that the animal brain is something so particular that it would be difficult to imagine reproducing it artificially. Turing brings the discussion to a close by stressing how he never believed one of his machines would actually reproduce the human brain and its physiology, but rather the thought processes; and he maintains that our technology would enable us to reproduce the thought processes mechanically without having to reproduce the whole brain as well.

Conclusion

Imitational creation is a theme that originated in myth and subsequently became part of the imagery of the modern world, where it gave rise to a plethora of artificial beings and robots on which many artistic creations, literature, plays and films have taken as their theme. Whenever science and technology appeared to be able to fulfil the human desire to create, the theme was taken up again. A major turning point for this mixture of imagination and technological achievements was marked by the birth of artificial intelligence, which seemed to provide the tools to make the old dream come true. The goal, however, was no longer a naive reproduction of man as a whole, but rather a detailed version of just one of his parts, the human mind. The theme of artificial creation on stage has been analyzed in three plays which, albeit in very different ways and perspectives, deal with some of the questions that inevitably come to mind whenever the possibility of creating thinking or even feeling machines, arises. *R.U.R. Rossum's Universal Robots* is a science fiction in which Karel Čapek foresees a society ruled by robots. Giuseppe O. Longo's *Il cervello nudo (The Naked Brain)*, is a drama that features a scientist who employs his knowledge to build machines only to abandon them on the verge of conscience in a very ambiguous territory. Here the beginnings of the individualization process turn out to be deeply cast in pain. *I Cinque di Cambridge (The Cambridge Five)*, as adapted from John L. Casti's own text is a conference-play about the possibility that machines could think. *Il Cervello Nudo* was developed for the scientific theatre and of the three plays mentioned above it is the one that most clearly shows how theatre and science can combine in the case where science is concerned with problems that deeply affect all human beings in their relationship to the world and in the responsibility humans take for their action. In other words, it is when science triggers off considerations that go beyond scientific matters to investigate the meaning of this world, of life and death that the theatre becomes once again the mirror of our own inner theatre known as 'conscience', where such matters are staged together with our thoughts, fantasies and dreams.

Bibliography

- [1] Umberto Artioli & Francesco Bartoli (edited by), *Il mito dell'automa. Teatro e macchine animate dall'antichità al '900*, ed. Artificio, Firenze, 1999
- [2] Silvana Barbacci, *Un caleidoscopio magico: la scienza a teatro*, Texts for a master in science communication by Silvana Barbacci, Trieste, 2001
- [3] Yuriy Castelfranchi - Oliviero Stock, *Macchine come noi*, Laterza, Roma – Bari, 2000
- [4] John L. Casti, *I cinque di Cambridge*, Raffaello Cortina Editore, Milano, 1998
- [5] Karel Čapek, *R.U.R & L'affare Macropulos*, Einaudi, Torino, 1971
- [6] Julian Jaynes, *Il crollo della mente bicamerale e l'origine della coscienza*, Adelphi, Milano, 1984
- [7] Giuseppe O. Longo, *Il cervello nudo*, La Contrada, Trieste, 1998
- [8] Giuseppe O. Longo, *Il nuovo Golem*, Laterza, Roma – Bari, 1998
- [9] Giuseppe O. Longo, “Quella macchina è incosciente”, in *La Rivista dei Libri*, n. 6, september 1991
- [10] Giuseppe O. Longo, “Matematica e arte”, in *La Rivista dei Libri*, n. 11, november 1992
- [11] Giuseppe O. Longo, “Ambiguità tra scienza e filosofia”, in *Nuova Civiltà delle macchine*, IX, n. 3-4, 1993
- [12] Giuseppe O. Longo, “Intelligenza della mente intelligenza del corpo”, in *Studium*, Edizioni Studium, Roma, 3 – 4 – 2000
- [13] Giuseppe O. Longo, *Il fuoco completo*, Moby Dick, Faenza, 2000

- [14] Giuseppe O. Longo, *Homo technologicus*, Meltemi, Roma, 2001
- [15] Giuseppe O. Longo, *Avvisi ai naviganti*, Moby Dick, Faenza, 2001
- [16] Mario G. Losano, *Storie di automi dalla Grecia classica alla Belle Epoque*, Einaudi, Torino, 1990
- [17] Gustav Meyrink, *Il Golem*, Bompiani, Milano, 1977
- [18] D.R. Hofstadter – D.C. Dennett, *L'io della mente*, Adelphi, Milano, 1985
- [19] Georges Ripka, “Una paura che viene da lontano”, in *Lettera Internazionale*, n. 63, 1° trimestre 2000
- [20] Teatro del Carretto (edited by), *I sogni dell'attore artificiale*, Lucca, Ottobre/Novembre 1986
- [21] Gianni Zanarini, “Le macchine pensanti”, foreword to *Il cervello nudo*, Giuseppe O. Longo, La Contrada, Trieste, 1998

Translated by Elisabetta Maurutto, Scuola Superiore di Lingue Moderne per Interpreti e Traduttori, Trieste, Italy.