Consiglio Nazionale delle Ricerche

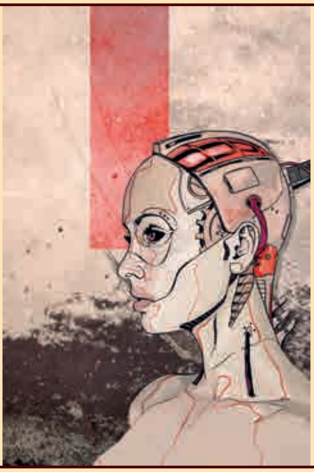
Wired Bodies

New Perspectives on the Machine-Organism Analogy

Editors

Nicole Dalia Cilia

Luca Tonetti



Filosofia e saperi - 9

Filosofia e saperi – 9

Sconfinamenti tra i saperi umanistici e le scienze della vita
Crossing borders between the humanities and the life sciences

Collana dell'Istituto per la storia del pensiero filosofico
e scientifico moderno (ISPF) del Consiglio Nazionale delle Ricerche

Book series of the Institute for the History of Philosophy
and Science in Modern Age, National Research Council, Italy

The book series “Filosofia e saperi”, active from 2009, has renewed in 2016 its scientific mission, adding the subtitle “Sconfinamenti tra i saperi umanistici e le scienze della vita / Crossing borders between the humanities and the life sciences”, enlarging its Scientific Committee and accepting texts also in English and in the other major European languages. Its scope is to promote research committed to a dynamical representation of the relationship between human sciences and life sciences and practices, and to stimulate new theoretical perspectives capable of supporting the communication and interaction between different disciplinary fields and thought styles.

The research fields addressed by the book series are:

- history and philosophy of the life sciences
- bioethics
- history of scientific concepts and metaphors
- public understanding of life sciences
- social and political history of life sciences
- history of scientific collections and museums
- art and life sciences iconography
- scientific practices and sites of knowledge production in life sciences

Contact for submissions: caianiello@ispf.cnr.it

www.edizioni.cnr.it
bookshop@cnr.it

P.le Aldo Moro 7
00185 Roma

ISBN 978 88 8080 236 5

I edizione: marzo 2017
Stampa Arti Grafiche Bruno - T. Greco

www.ispf.cnr.it

This volume is funded by the PhD program in Philosophy, Department of Philosophy, Sapienza University of Rome

Wired Bodies

New Perspectives on the Machine-Organism Analogy

Editors
Nicole Dalia Cilia
Luca Tonetti

Contributors
Francesco Bianchini
Elisa Binda
Giuseppe Boccignone
Nicole Dalia Cilia
Francesco De Bei
Mattia Della Rocca
Simone Guidi
Fiorenza Lupi
Alessandra Passariello
Francesco Restuccia
Luca Tonetti
Valentina Trombetta

Table of Contents

Preface

Elena Gagliasso

The reading group "Evolution & Cognition" was established in March 2012 at the Department of Philosophy of Sapienza University in Rome. The group met once a week to review a scientific paper, chosen from a list proposed by its members. The first members were, in particular, PhD students in Philosophy of Science (Laura Desirée Di Paolo, Fabio Sterpetti, Andrea Raimondi, Flavia Fabris, Diego Antonio De Simone), but it was immediately extended to PhD students and researchers from other disciplinary fields, such as Moral Philosophy (Eleonora Severini) and even Palaeoanthropology (Fabio Di Vincenzo). Despite its intense schedule, the group quickly started to attract both undergraduate and master students (Jacopo D'Alonzo, Giuliana Pulvirenti, Valeria Di Giovannandrea, and Ivan D'Annibale). More recently, it has welcomed new PhD students in cognitive science (Nicole Dalia Cilia), history of science (David Ceccarelli, Luca Tonetti), and semiotics (Massimiliano Napoli).

The group "Evolution & Cognition" has born spontaneously, and has rapidly developed in a multidisciplinary direction, inspired by the pluralistic approach in Philosophy of the Life Sciences. Its interest was mainly focused on concepts related to the latest advancements in science and contemporary philosophy: in the theory of evolution and cognitive studies, in philosophy of biology and philosophy of mind, epigenetics, behavioural studies on non-human primates, niche-construction, neurobiology, the relationship between morality and evolutionary biology, human evolution, and particularly the epistemological issues arising from the clash between realist and constructivist explanations in naturalized approaches.

The activities of the earlier group encouraged its members to share knowledge and familiarize, often building long lasting friendships. How to make available to others the experience, limited in time, of such a group that existed for pure passion? One of the main difficulties was the annual rotation of PhD students. The academic landscape in our country was then (and still is!) experiencing a dire recession, whose most direct effect on the group was the diaspora of young researchers, many of which have now taken up positions in different universities throughout Europe (Göttingen, Exeter, Nottingham, Berlin, and Paris).

Since the very start, the founding members of "Evolution & Cognition" have felt the need to give the group an institutional structure, so as to guarantee continuity in time. Thanks to the Doctoral College, coordinated by Professor Piergiorgio Donatelli, in September 2013 the reading group was transformed into a Permanent Seminar, associated to the PhD Program in Philosophy, with Roberto Cordeschi and me as institutional supervisors. Visiting lecturers, professors and young researchers, and graduate students from other faculties, as well as researchers and internal PhD students, were the invited speakers.

Roberto was then already in the last year of his life: he had been the mentor of several of the founding members of the group. In particular, his last student, Nicole Dalia Cilia, ultimately inherited the responsibility of organizing the Seminar. With Roberto Cordeschi, and then alone, I have been a supporter rather than a supervisor: the autonomy of decisions remained and remains with the students.

Since the end of 2013, it was decided that two organizers would take care of the Permanent Seminar alternating every two years: first, the founding members Laura Desirée Di Paolo and Diego Antonio De Simone; then, as the Seminar broadened its scope and changed its name in "Ecoevoluzione e Cognizione" (with the abbreviation ECOEVOCOG), Nicole Dalia Cilia and Luca Tonetti. For the year 2017, Alessandra Passariello and Stefano Pilotto will be the new organizers.

Retrospectively, the Reading Group has represented one of those germinal moments achieved thanks to a shared concert of interests and that becomes formative for its participants, acting as a *reflexive feedback*. Among young researchers, this is necessarily «quelque chose d'insolite, ou d'insolent» in the words of Pierre Bourdieu (Bourdieu 1982, p. 35).

Despite its transformation into a more institutionalized form, the Permanent Seminar has increased the contribution of external voices, without giving up its original freedom of thinking. Just like the Reading Group, the same idea was at play: considering the philosophy of science and the history of life sciences as a moment of critical, analytical and productive reflection, focused on the new insights from the contemporary scientific topics chosen each year.

Indeed, the practice and the experience within the Seminar became, for all those concerned, a place where its members can distinguish the «propre» from the «non-propre». This has been possible because this group has achieved the rare intent of a 'true' seminar: in Michel de Certeau's words, «ouvr[ir] une porte de sortie et de rentrée» (De Certeau 1978, p. 177), offering the means to distance oneself from the original landscape of one's own ideas in order to «y retourner autrement» (*ibidem*). Such a seminar was, and still is today, aimed at fostering critical research on the philosophical issues selected each year, going «contre sa formation autant qu'avec sa formation» (Bourdieu 1982, p. 6) and trying to avert «ce principe systématique d'erreur qu'est la tentation de la vision souveraine» (*ibidem*, p. 8).

Indeed, I think that this volume faithfully conveys to the reader the idea of this "journey" through different paths that are at the same time explanatory, justificatory, constructive as well as critical, while avoiding the presumption of completeness.

For 2015-16, the organizers Nicole Dalia Cilia and Luca Tonetti chose as the main topic the classical analogy between Machine and Organism, providing PhD students and professors the necessary background for the debate. The different issues raised by this powerful analogy are explored in the three sections of this book, from a historical and epistemological point of view. This Seminar was particularly endorsed by the "Interuniversitary Research Centre on Epistemology and History of Life Sciences" (Resviva). At a later time, the issues addressed have been also the subject of an entire section (*Brain and Behaviours in Ecological Frameworks*) of the International Workshop for the 20th Anniversary of Resviva "*Sliding Doors. Prediction and Contingency in Biosciences*" (February 2-4, 2017).

Unfortunately, the papers collected in this volume cannot recreate the atmosphere of the discussion and of the general interaction "in vivo". However, they show the synergy between different disciplines, biology, medicine, neuroscience, and artificial intelligence, all addressing the problem of the machine-organism analogy.

They fruitfully address the transition from the mechanical to the simulative approach, as well as the development over time of their explorative methodologies: from mechanistic and "atomizing" to informational and virtualizing methodologies. Moreover, they allow the reader to retrace the historical reasons and the extra-scientific factors behind this influential analogy that, as an "inexhaustible engine", passes through about three centuries of Western science and culture.

One long argument that, as the reader will acknowledge, is at the same time philosophical, scientific, cultural, and historical; that allows grasping each time a point of view, a perspective, but that is never a mere element of a summation, rather a node in a network, always open to further changes.

References

Bourdieu P. 1982, *Leçon sur la leçon*, Ed. de Minuit, Paris.
De Certeau M. 1977, *Qu'est-ce qu'un séminaire?*, in «Esprit», 22/23(11-12), pp. 176-181.

Introduction

Nicole Dalia Cilia
Luca Tonetti

In the 1920s the German physician Fritz Kahn (1888-1968) published a series of illustrated popular science books entitled "Das Leben des Menschen", devoted to the description of human physiology (Borck 2007; Eilers 2015). The main aim of this work was to represent bodily processes, as opposed to anatomical structures, by means of the machine-organism analogy. One of its most famous plates (Figure 1), – known as the "Man as Industrial Palace" (*Der Mensch als Industriepalast*) – shows the functioning of the upper part of the human body as a set of separated machines, working in a sort of production line. It thus brings together, in a peculiar way, medicine and technology. The "human factory", in fact, is composed of steam engines, a ventilation system, pistons, a chain of conveyer belts, tanks – each of them representing a single function of the body.

Figure 1. *Das Leben des Menschen* in the Arthur and Fritz Kahn Collection. Source book page: https://archive.org/stream/arthurfritzkahn_04_reel04/#page/n220/mode/1up

Kahn's mechanical imagery inspired other artists, among them the Scottish sculptor Eduardo Paolozzi (1924-2005), who in 1970 presented the *Conditional Probability Machine,* a portfolio of 24 etchings, divided in four sections – "Secrets of Life – The Human Machine and How it Works", "Manikins for Destruction", "Pages from the Aerospace Medical Library", and "From Genot to Unimate" –, whose men-machine images were taken from the illustrations of popular science journals and books. In particular, the first and fourth sections were devoted, respectively, to the functioning of human body and to the representation of different kinds of robots, humanoids too.

The machine-organism analogy (hereafter MOA) has played such a significant role in the history of Western philosophy and science that its influence on a variety of aspects of popular culture – ideas, attitudes, images, media but also language – and its presence in everyday life of contemporary societies can not be exaggerated. Although it sometimes seems to be applied naively, the analogy hides complex epistemological issues, concerning the status of both organism and machine, as well as the nature of their interaction.

As far as history is concerned, mechanical analogies can be dated back to Aristotle. Nevertheless, our goal here is not to retrace in detail their historical evolution. It is worth stressing, instead, that pre-modern forms of machine-organism analogies do not always build on a mechanistic conception of living beings. In other words, mechanical analogies do not necessarily imply some form of mechanistic reductionism. This is the case, for example, of another artist, one icon of the Western tradition, Leonardo da Vinci. In his *Anatomical Drawings* (preserved in the Royal Library at Windsor Castle) he argued that nature had provided living beings with «strumenti macchinali», without which some processes, such as movement, would not be possible[1]. This implies that the "dissection" of machines into their constituent parts could well serve the study and description of bodily processes; however, since Leonardo accepts the traditional notion of the soul, the identification of mechanical structures does not mean that the human body in itself is a machine. As the artists' use of the analogy well shows, MOA is fraught with difficulties and can not be easily reduced to simple patterns.

Wired bodies. New perspectives on the machine-organism analogy provides the reader with some suggestions on these issues, coming from different perspectives. Its main contention is that this analogy is not static, but dynamic: it strictly depends, in fact, on the way both categories, "machine" and "organ-

[1] Windsor, Royal Library, K/P 153r.

ism", have developed over the centuries. It also depends on the interaction with other and diverse scientific and philosophical assumptions and models: the plasticity of MOA is apparent in the many different points addressed in the contributions to this volume. More specifically, the purpose of this book is the description of 1) the development of a "mechanistic" framework in medicine and biology; 2) the methodological issues underlying the use of 'simulation' in cognitive science; and 3) the interaction between humans and machines according to 20th-century epistemology.

The development of a 'mechanistic' framework in medicine and philosophy

In the 17th century, a variety of mechanistic models comes to the fore, advanced from different perspectives and serving different purposes. In the field of anatomy, physiology and medicine, many of them intersect and productively interact with other models of the body, including those constructed by iatrochemistry. In fact, chemical knowledge and practice did offer viable solutions to a number of problems affecting the early-modern ultra-mechanistic machine of the body, among them prominently those concerning animation, thought and nervous processes, and movement. This hybridization and cross-fertilization has been neglected, if not completely forgotten, by a historiography focusing on neat distinctions and categories (see e.g. Dijksterhuis 1961; Debus 1977).

As regards philosophy, the novelty of Descartes' view is not in MOA analogy itself, which already existed, but in its dependence, in his works, on a more general philosophical perspective. This aims at reducing the physical world to geometric properties, such as extension and motion, as a consequence of the dualism of substances – mind vs. matter. Organic processes of the body can thus no longer be explained by recourse to the action of a soul; consequently, death «never comes to pass by reason of the soul, but only because some one of the principal parts of the body decays» (Descartes 1985, p. 329). The implications of this view in ontology (substance dualism), in psychology (the status of the soul and the mind-body problem) and in physiology (the action of animal spirits and the function of the pineal gland), along with its even more radical developments in the post-Cartesian debate (e.g. see Lamettrie and his *L'Homme machine*), are well known and have been recently reassessed (Bitbol-Hespériès 1990; Des Chene 2001; Aucante 2006; Caps 2010; Allocca 2012; Scribano 2015).

However, much has obviously changed since Descartes, and today MOA requires a more detailed and thorough analysis. Whenever we say that, in some sense, "humans *are* machines" or "organisms can be explained *mechanistically*", we assert something radically different from what was originally claimed by Descartes. What is the real object of this analogy: *organisms* as a whole, their *parts* or, rather, bodily *functions*? How can the machine serve as a model for interpreting the biological phenomena, the cognitive processes, or more broadly the social and cultural transformations of the relations between individuals, and between individuals and the environments in which they live?

The history of early modern medicine is interesting here, in that it may help clarifying some implications of the philosophical debate. Medicine shows, in particular, that MOA can be properly applied to the analysis of normal as well as to morbid states of the body. This is evident in the rise of the so-called "mechanistic anatomy", which, pursuing Cartesian mechanism, attempted to explain in mechanistic terms the structures and the functions of the body parts, both in healthy and diseased conditions. In recent studies, Bertoloni Meli explains how the use of mechanical models in anatomy might have influenced the understanding of human body; however, it could also have allowed physicians to treat morbid states "mechanistically". In other words, physicians could build machines for clinical purposes, i.e. they could use physical artefacts or mechanical devices to reproduce morbid operations, in order to grasp the natural laws underlying them and find a suitable treatment (Bertoloni Meli 2007, 2008, 2011, 2012, 2016; Bertoloni Meli, Wilkin 2008).

Moreover, as Bertoloni Meli himself recognises, the concept of "machine" is very broad, since it includes traditional machines (like clocks, for example) as well as hybrid ones that are based on chemical processes. This is only part of a more general problem concerning the distinction between the two medical doctrines, iatrochemistry and iatromechanics: iatrochemistry sees the biological and physiological phenomena in chemical terms, interpreting them as processes of "fermentation" and "effervescence"; iatromechanics, conversely, applies the physical laws of mechanics and statics to bodily operations. However, as already mentioned, this distinction does not reflect the complexity of early modern medicine. A leading representative of iatromechanics like Giovanni Alfonso Borelli, for example, adopts chemical explanations next to purely mechanical categories to explain the movements of the body or generation.

The first two papers in the book, by Guidi and Tonetti, aim at analysing forms of mechanism alternative to Descartes' celebrate and well-known model, in order to show that the notion of "mechanical philosophy" and of "machine" are not unitary, and by no means simply reducible to Cartesianism alone. Simone Guidi examines Cureau de La Chambre's battle against Descartes' mechanical materialism, questioning the structural relation of mechanicism to materialism by showing that it can be reconciled, rather surprisingly, with an updated version of ancient hylomorphism. Luca Tonetti explores an alternative medical version of mechanicism by describing Giorgio Baglivi's mechanistic pathology. Although he is traditionally considered a leading representative of the Italian "iatromechanics", Baglivi realized that diseases can not be entirely reduced to physics, since the morbid states of the body are completely different from machines malfunctioning. By taking into account von Bertalanffy's early work, Alessandra Passariello shows that the major arguments of early 20th-century organicism against a mechanized notion of living entities were both absorbed and overcome by the upcoming cybernetic approach. Organicist arguments thus prepared the ground for a substantial reconfiguration of the organism vs machine dichotomy into a more complex one, namely, organism-machine vs non-living complex dynamical systems. Finally Mattia Della Rocca reflects on the concept of simulation shared by recent European initiatives like the Human Brain Project. In particular, he focuses on how the adoption of an engineering perspective and the action of extra-scientific factors, like research policies and procedures, influence current neuroscience.

Understanding by building. The use of simulation in cognitive science

Vittorio Somenzi's paper "Men and Machines", presented at the Italian Congress of Philosophy held in Pisa in April 1967, was one of the first Italian contributions on the relationship between organism and machine (Somenzi 2011)[2]. Somenzi claims that Descartes, Kepler, Boyle and Leibniz had already recognized the existence of machines aimed at reproducing not only muscular, but also mental work, as exemplified by the working of clocks. In the latter case, the clock is not just a tool, but a real machine, which possesses autonomy and is able to give a feedback by returning information which was previously entered. "Autonomy", in particular, represents a turning point with

[2] All quotations from Somenzi are translated by the authors.

respect to 17th-century machines, which released instantaneously the energy stored in muscles, because it allows the mechanical devices to hide their source of energy (*ibidem*, p. 131). The "hidden engine" has all the information «that the automaton would then have performed [...] in different kinds of outputs, such as gestures, sounds, writing, and so on» (*ibidem*), that is the process of mental activity simulated by the robot. Other machines performing mental works, rather than muscular efforts, were mechanical calculators or programmable machines like automatic looms, which were developed between the 17th and 18th centuries.

Between the 1830s and 1870s, the English mathematician Charles Babbage (1791-1871) tried to combine in his *Analytical Engine* (Figure 2) the concept of "program", which is implicit in automatic looms, with the concept of the calculating mechanism of the "pascaline", a calculating machine previously invented by the French mathematician and philosopher Blaise Pascal.

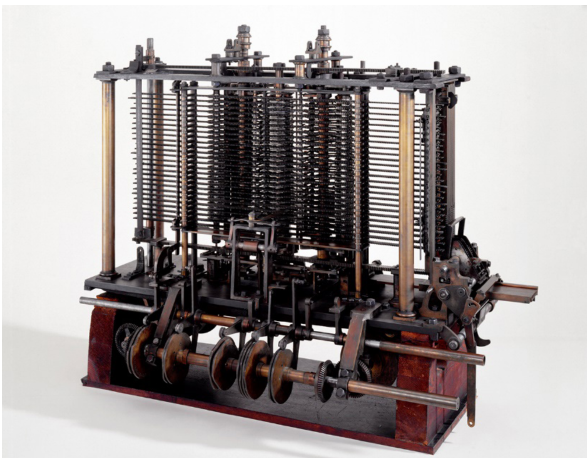


Figure 2. Babbage's Analytical Engine.

Thus the idea was born of a general-purpose calculator, i.e. of a machine able to solve very different problems depending on the program implemented therein. After almost a century, Alan Turing (1912-1954), in his famous paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*" (1936), generalized this idea in a more abstract form. Turing's machines can imitate human deductive reasoning but, according to Somenzi, also the «operations characterizing the inductive processes and any other mental process, as long as these operations are defined uniquely with a finite number of words» (*ibidem*, p. 132) (for more details, see Tamburrini 2002, pp. 29 ff).

At the engineering level, the predecessors of the first generation computers were mechanical or electromechanical devices. The basic components were spring-loaded relays whose operations closely resembled those of ordinary switches: they opened and closed a circuit in response to an electrical signal (Cordeschi 1991). With the so-called "electronic revolution" the context changes. Since the 1960s, «the construction of integrated circuits at steadily rising levels of miniaturization and speed, the so-called chips, has led to previously inconceivable results» (*ibidem*, p. 315). However, these technological advances have not produced any change in the structure of programs: «even neurons are triggered or not, a nervous impulse travels or not» (*ibidem*). This is the meaning of McCulloch's brain-computer analogy, which inspired the first modern comparisons between "electronic brain" and "biological brain" (*ibidem*).

The American logician Walter Harry Pitts (1923-1969) contributed to the application of mathematics to the biological sciences. With McCulloch, he proposed the first mathematical model of a neural network based on propositional logic, aimed at reproducing the same over-all properties of synapses (for the McCulloch–Pitts artificial neuron, see McCulloch, Pitts 1943):

> The all-or-none character of the discharge of the neurons is precisely analogous to the single choice made in determining a digit on the binary scale, which more than one of us had already contemplated as the most satisfactory basis of computating-machine design. The synapse is nothing but a mechanism for determining whether a certain combination of outputs from other selected elements will or will not acts as an adequate stimulus for the discharge of the next elements, and must have its precise analogue in the computing machine. The problem of interpreting the nature and varieties of memory in the animal has its parallel in the problem of constructing artificial memories for the machine (Wiener 2013, p. 14).

McCulloch-Pitts artificial neuron inspired Frank Rosenblatt's work on the "perceptron" (Figure 3), which is defined as a

> hypothetical nervous system, or machine […] designed to illustrate some of the fundamental properties of intelligent systems in general, without becoming

too deeply enmeshed in the special, and frequently unknown, conditions which hold for particular biological organisms (Rosenblatt 1958, p. 387).

The perceptron allows a rudimentary form of adaptation or "learning": the network of artificial neurons receives excitatory impulses from other neurons, and the weight of each impulse can be adjusted during the "learning period".

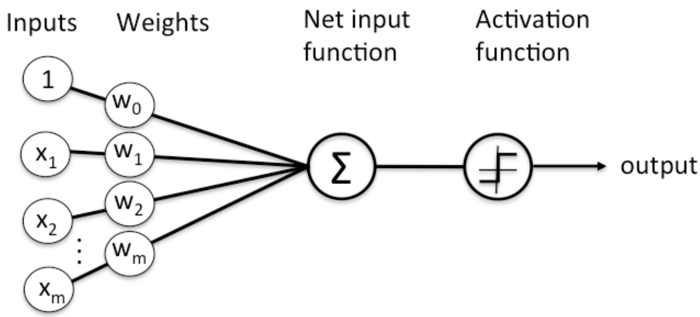


Figure 3. Organization of a perceptron. Source: Rosenblatt 1958.

Somenzi claims that the McCulloch-Pitts model has shown that the human brain is similar to the Turing machine and that «any Turing machine could be realized by means of neurons of which it is composed» (Somenzi 1967, p. 132). In this way, Somenzi offers a definition of "machine", namely, «a system able to assume one among different states, a list that associates each of these states to a particular action, and the rules to switch from one state to another» (*ibidem*). This definition shows how the machine design has developed over time, from traditional mechanical devices such as watches or looms, towards virtual and computerized machines. The use of these kind of machines is now frequently based on «mathematical models for the explanation of human mental activity, as well as for the design of more efficient artificial substitutes for our brain» (*ibidem*, p. 133).

Building on Somenzi's interpretation, the second section of this volume deals with the problem of artificial simulations in current cognitive science. In 1943, in one of the most important meetings in the history of cybernetics, engineers, physiologists and mathematicians produced the *manifesto* of cybernetics (Rosenblueth, Wiener, Bigelow 1943). Research in this field is based on the possibility of considering machine and brain as functionally equivalent. Since the 1940s, in order to reproduce and investigate the mechanisms of cognitive functions, cognitive scientists have followed a methodology known as the synthetic method (Cordeschi 2002). Through this method, the machine can be used for testing theories, since "mechanical organism (or machine)" and "biological organism" (or proper body) seem to share some essential characteristics of the investigated phenomenon, revealing a common functional organization beyond the different physical structures (Cordeschi 2008). Therefore, the goal of the synthetic method is not to reproduce cognitive functions but to test their "mechanisms" on the basis of which machines are built. This is possible by comparing the behaviour of the machine with that of humans. Obviously, many difficulties arise from these assumptions, such as the problem of functionalism, sufficiency test and multiple realizability.

Nicole Dalia Cilia analyses the difference between *model-oriented* simulation and *data-oriented* simulation (as in the synthetic method) arguing that it is possible to integrate the data-oriented simulation in the first one. Giuseppe Boccignone applies this methodology to an issue, emotions, until now considered as an exclusive domain of philosophy or psychology. While addressing the same question, Valentina Trombetta examines the role of emotions and intrinsic motivations through an enactive and dynamical framework. Finally, Francesco Bianchini discusses the relationship between evolutionary biology and AI.

Bodily boundaries and beyond: how do machines extend the bodies and their world?

Contemporary science poses new challenges: modelling mechanical devices can be difficult, but managing the interaction between humans and machines may pose further problems. The third section addresses this issue by adopting as a background recent philosophical trends and debates. Technology has become a tool for acquiring knowledge as much as a practical support, as apparent in the production of artificial prostheses in medicine, in rehabilitation equipment, in the assessment for children with learning disability, in home automation or in tools for large-scale control. The world we live in, being largely technologized, is a human product that, in some sense, "extends" and enhances our capabilities. Therefore, the third section explores the ways technology can extend bodily, mental and evolutionary boundaries, arguing that our body is an extension of our brain, as machines are extensions of our bodies. In particular, Lupi and Binda investigate the role of technical objects, interpreted as extensions of our body. Fiorenza Lupi focuses her paper on Georges Canguilhem's philosophy of technique, overturning the traditional relation between machine and organism. The question here is: do machines show adaptation, learning and interaction with the environment? Elisa Binda addresses the relation among technical objects, organisms and environment, according to the analysis by Gilbert Simondon, one of Canguilhem's disciples. Francesco Restuccia shows that Wiener and Benjamin, although coming from different backgrounds, share the same interest about the interaction between human beings and machines, and the same concern about the impact of technology on contemporary societies. However, the notion of "extension" is supported also by embodied cognition (Shapiro 2012) and dynamic psychology (Dazzi, De Coro 2001), which reassess the mutual relationship between organisms and their environments, in order to provide a more *situated* conception of cognitive processes. Francesco De Bei exploits the evolutionary hypotheses developed in recent years by Michael Tomasello to propose a new interpretation of human-machine symbiosis, also taking into account the impact that an environment increasingly *saturated* of technology could have on mental functioning.

Waiting for the day when we can all «journey to a world where robots dream and desire»[3], we do hope that this volume will continue the fruitful discussions which began as a seminar, and will suggest to the reader new perspectives on the analogy between machine and organism.

Acknowledgments

This book is the result of a series of meetings on "Machine and organism", organised by the Permanent Seminar "ECOEVOCOG" within the activities of the PhD Program in Philosophy. The seminar was held from March to June 2016 in Villa Mirafiori, at the Department of Philosophy, Sapienza University of Rome, and saw the participation of scholars coming from coming from a variety of research institutions, such as the University of Rome

[3] Tagline of *A.I. Artificial Intelligence*, a 2001 science-fiction film directed by Steven Spielberg.

Tor Vergata, the University of Pisa, the University of Milan, the Nuova Accademia di Belle Arti (Milan), the University of Bologna, the University of Padova, the University of Freiburg. These events were kindly endorsed by the *Interuniversity Research Centre on Epistemology and History of Life Sciences* (ResViva) and sponsored by numerous academic societies: the *Italian Association for Cognitive Sciences* (AISC), the *Italian Association for Artificial Intelligence* (AI*IA), the *Circolo Bateson*, and the *Italian Association of Systemic Epistemology and Methodology* (AIEMS).

We wish to thank the Director of the PhD Program in Philosophy at Sapienza University, Prof. Piergiorgio Donatelli, for having encouraged our project. In particular, we also wholeheartedly thank Prof. Elena Gagliasso, for her guidance and intellectual stimulation.

However, this book would not have been possible without the patient and constant support of Silvia Caianiello and Maria Conforti, who accepted *Wired bodies* within the series "Filosofia e Saperi" of CNR Edizioni and assisted us in the last phase of editing process, by revising and checking all papers. Their assistance has been decisive for the overall success of this work.

Selected list of references

Allocca N. 2012, *Cartesio e il corpo della mente*, Aracne, Roma.

Aucante V. 2006, *La philosophie médicale de Descartes*, PUF, Paris.

Bertoloni Meli D. 2007, *Mechanistic Pathology and Therapy in the Medical Assayer of Marcello Malpighi*, in «Medical History», 51, pp. 165-180.

Bertoloni Meli D. 2008, *The Collaboration between Anatomists and Mathematicians in the mid-17th Century. With a Study of Images as Experiments and Galileo's Role in Steno's Myology*, in «Early Science and Medicine», 13(6), pp. 665-709.

Bertoloni Meli D. 2011, *Mechanism, Experiment, Disease. Marcello Malpighi and Seventeenth-Century Anatomy*, Johns Hopkins University Press, Baltimore.

Bertoloni Meli D. 2012, *Machines of the Body Between Anatomy and Pathology*, in *Modèle Métaphore Machine Merveille*, ed. by A. Gaillard, J.Y. Goffi, B. Roukhomovsky, S. Roux, Presses Universitaires de Bordeaux, Bordeaux, pp. 53-68.

Bartoloni Meli D. 2016, *Machines of the Body in the Seventeenth Century*, in *Early Modern Medicine and Natural Philosophy*, ed. by P. Distelzweig, B. Goldberg, E.R. Ragland, Springer, Dordrecht-London, pp. 91-116.

Bertoloni Meli D., Wilson R., eds. 2008, *Special issue: Observation and Experiment in Mechanistic Anatomy*, in «Early Science and Medicine», 13(6), pp. 531-709.
Bitbol-Hespériès A. 1990, *Le principe de vie chez Descartes*, Vrin, Paris.
Borck C. 2007, *Communicating the Modern Body: Fritz Kahn Popular Images of Human Physiology as an Industrialized World*, in «Canadian Journal of Communication», 35, pp. 495-520.
Caps G. 2010, *Les medecins cartesiens: héritage et diffusion de la représentation mécaniste du corps humain (1646-1696)*, Olms Verlag, Hildesheim-.
Cordeschi R. 1991, *Brain, Mind and Computers*, in *The Enchanted Loom: Chapters in the History of Neuroscience*, ed. by P. Corsi, Oxford University Press, Oxford, pp. 315-320.
Cordeschi R. 2002, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Kluwer, Dordrecht.
Cordeschi R. 2008, *Steps Toward the Synthetic Method: Symbolic Information Processing and Self-Organizing Systems in Early Artificial Intelligence*, in *The Mechanical Mind in History*, ed. by P. Husbands, O. Holland, M. Wheeler, MIT Press, Cambridge (MA), pp. 219-258.
Dazzi N., De Coro A. 2001, *Psicologia dinamica. Le teorie cliniche*, Laterza, Roma-Bari.
Debus A.G. 1977, *The Chemical Philosophy: Paracelsian Science and Medecine in the Sixteenth and Seventeenth Century,* Science History Publications, New York.
Descartes R. 1985, *The Passions of the Soul* (1649), in *The philosophical writings of Descartes*, ed. by J. Cottingham, N. Stoothoff, D. Murdoch, vol. I, Cambridge University Press, Cambridge.
Des Chene D. 2001, *Spirits & Clocks. Machine & Organism in Descartes,* Cornell University Press, Ithaca (NY)-London.
Dijksterhuis E.J. 1961, *The Mechanization of the World Picture*, tr. by C. Dikshoorn, Oxford University Press, New York.
Eilers M. 2015, *Fritz Kahns „Das Leben des Menschen". Zur Produktion und Transkription eines populären Werks*, in «NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin», 23(1), pp. 1-31.
La Mettrie J. Offray de 1747, *L'Homme machine*, Leyden.
McCulloch W., Pitts W. 1943, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in «Bulletin of Mathematical Biophysics», 5, pp. 115-133.
Rosenblatt F. 1958, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, in «Psychological Review», 65(6), pp. 386-408.

Rosenblueth A., Wiener N., Bigelow J. 1943, *Behaviour, Purpose, and Teleology*, in «Philosophy of Science», 10, pp. 18-24.
Scribano E. 2015, *Macchine con la mente. Fisiologia e metafisica tra Cartesio e Spinoza*, Carocci, Roma.
Shapiro L. 2011, *Embodied Cognition*, Routledge, London-New York.
Somenzi V. 2011, *Uomini e macchine*, in *Vittorio Somenzi. 1918-2003. Antologia e testimonianze*, ed. by B. Continenza, G. Corbellini, R. Cordeschi, E. Gagliasso, C. Morabito, M. Stanzione, Fondazione Banca Agricola Mantovana, Mantova, pp. 130-139.
Tamburrini G. 2002, *I matematici e le macchine intelligenti*, Mondadori, Milano.
Turing A. 1936, *On computable Numbers, with an Application to the* Entscheidungsproblem, in «Proceedings of the London Mathematical Society», 42, pp. 230-265.

I

The Development of a "Mechanistic" Framework in Medicine and Philosophy

Mechanism Prehistory and the Strange Case of Cureau de La Chambre

Simone Guidi

What does the word "mechanism" mean? Introducing his masterful *The Mechanization of the World Picture*, Dijksterhuis admitted to not having a final answer (Dijksterhuis 1961, p. 4 ff). His work dealt with the entire history of natural philosophy, concluding that the mechanization of the world «meant the introduction of a description of nature with the aid of the mathematical concepts of classical mechanics; it marks the beginning of the mathematization of science, which continues […] in the twentieth century» (*ibidem*, p. 501).

Dijksterhuis' emphasis lays on mathematics, although it is evident how such a definition pays the price for placing mechanism's roots so far back in time. Modern science would have been living for centuries in the trail of classical philosophy, and most of the philosophers we call "mechanists" could not be considered more than followers of the ancients' footsteps. Furthermore, neither Descartes nor Gassendi applied mathematics to physics, while Galileo – who can hardly be defined a "mechanist" (Garber 2013, pp. 13-15) – could not be excluded from this category. For mechanism, every physical phenomenon, including *force*, can be entirely reduced to spatial relations, and Galileo's physics leaves many fields out of the reduction to space.

Mechanism is a "theory of everything", whose attempt is to reduce every science to the same, intelligible principles. Consistent with this perspective is a simpler definition, as the one we find in Westfall's *The Construction of Modern Science* (Westfall 1977). For him, the mark of "mechanical" philosophy is the overcoming of Renaissance natural philosophy (*ibidem*, p. 30). The core of Renaissance philosophy being the belief in a universal, *spontaneous* animation of matter, the crucial passage toward the rising of mechanism would be represented by the homogenization of matter under a general notion of "body".

This is the reason why the very notion of material body should be specified. While it has represented for centuries a mere potentiality (Aristotle), a part of a hylomorphic composite (Aquinas) or a metaphysical substance (Avicebron, the Franciscans), in the 17th century it starts to coincide with a geometrical *quantitas* (extension) that needs matter as a *substratum* for its ontological independence. Such a "general theory" of extended bodies perfectly matches Ayer's definition of mechanism as a system in which «the laws of physics can be explained […] by being deduced from the attributes possessed essentially by all bodies *qua* bodies» (Ayers 1988); but it needs extension really more than it needs matter.

We might say that for early modern "mechanists" the identification of body, extension and matter – as well as the possibility of reconnecting every physical phenomenon to its uniformity and substantiality – still finds its roots in a *metaphysica generalis* (if not in a *specialis*). This seems the reason why it cannot be disconnected from a revolution in late Renaissance Aristotelianism, especially in the way in which "form" starts to be an *efficient* more than a *formal* cause of bodies.

In Carraud's work (Carraud 2002), we discover how, especially in Suárez, Aristotelian physics becomes a theory of efficient causation. The late Aristotelian world is a hierarchical apparatus of efficient causes, in which the 'first' One pre-determines nature as an independent whole of 'secondary' causes, leaving it to its inner execution and rejecting any role for occult qualities (Courtine 1990; Des Chene 1996; Coujou 1999; Wadell 2015).

In this context, Jesuit physicists dealt with a different hylomorphism too. As stressed especially by Courtine (Courtine 1990) and Coujou (Coujou 1999), Suárezism reduces his world to the 'simple' ontology of the *conceptus obiectivus* (Forlivesi, Boulnois 2002; Esposito 2014). Such a revolution also includes matter, to which Suárez acknowledges a subsistence that is independent from the form. This real-potentiality of matter leaves to form nothing more than an efficient, automatic, shaping action. Thus, for early modern hylomorphism, form is more a *disposition* (and even a pre-disposition) of matter than its very being, as, by contrast, it was assumed to be in Aquinas' hylomorphism. The efficient action of the 'local' causes provides a complete, essential explanation of phenomena. In some cases, it also includes the biological functions of the body, as in Gomez-Pereira's *Antoniana Margarita* (Sanhueza 1997), the most striking antecedent of Descartes's *bêtes machines*, discussed and criticized already in its very Aristotelian context (Suárez 1635). So, can we already label this early modern hylomorphic "automatism" as a "mechanism"?

Addressing "mechanism" from a historiographical perspective, Roux (Roux, 1996, 2013) and Garber (Garber 2013) highlighted how the expression «mechanical philosophy» does not appear before Boyle. For him, "mechanism" lies in one «Universal Matter common to all Bodies», that is, «a Substance extended, divisible, and impenetrable» into which variety is continuously introduced by «Motion». Matter has for Boyle «two Attributes»: «its own Magnitude, or rather *Size*, and its own *Figure* or *Shape*» (Boyle 1999-2000, vol. 5, pp. 305-307). What is new with respect to Descartes, Hobbes or Gassendi? Looking at Boyle's mechanism, we should focus on its *epistemology* rather than on its theorical conclusions. While the early 17th century's "mechanists" commonly accepted the belief that physics can be reduced to matter's configuration, Boyle's is an *experimental* science in which the experimental reality of matter makes metaphysical explanations completely useless. And this makes his mechanism completely different from the philosophical model developed by Descartes, Hobbes and Gassendi, more interested in finding analogical "general" explanations rather than in directly observing nature.

Significantly, Boyle enrolls them in the list of his predecessors, as the «excellent authors», who have in common «their opposition to Aristotelian physics» (*ibidem*, vol. 5, p. 295). Like Descartes' and Hobbes', Boyle's science lies in the refutation of Aristotelian *epistemology*, as well as in the rejection of a way of reasoning that still belongs to the early 17th century "mechanists". This is the reason why, in a historical perspective, it is hard to find an actual "break" between a "conceptual" mechanism that is still Aristotelian even while it replaces first philosophy with a new, science-friendly, metaphysics and with a Boylean experimental mechanism that flows from the breaches that this new metaphysics disclosed. Hence, Descartes, Hobbes and Gassendi cannot be considered «different variants of a common paradigm, but elements of a pre-paradigmatic stage in the development of the mechanical philosophy, part of its pre-history rather than of its history» (Garber 2013, p. 26); and a crucial consequence of this assumption is that such a "still-not-Boylean" mechanism includes a variety of philosophical perspectives, often in conflict with each other.

The plexus of prehistori(ographi)c mechanism brings together several philosophical schools that, at the end of the 16th century, give birth to an international debate. It provides a simpler epistemological framework for Boylean and Newtonian physics. An exploration in late scholastic physics and psychology – like Edward Grant's – will persuade us of a crucial fact: such an overlapping of perspectives has been permitted by the inner transformation of Aristotelian physics and in particular by the crisis of Dominican hylomorphism. For many years, scholastic philosophy had imprisoned any reality of matter in the shaping hands of form, and this had allowed the persistence of a metaphysical *medium* (the form itself) between logic and physical reality. Nature showed a logical behaviour as any natural motion was mediated by form, and conversely any natural motion could be described in logical terms, because of its inner formality.

The slow recognition of the reality of matter that, also within the Aristotelian tradition, starts at the end of the 16th century, directly places physical phenomena in front of a non-metaphysical logic, an ontology implicitly tending to nominalism like the one by Suárez (Esposito 2014, pp. 135-136). However, what remains of Aristotelianism when nature is completely described in terms of act and potency, efficient causes and actually-nominal essences? It is apparent that, starting at least with Jesuits' eclecticism, Aristotelianism addresses nature rather from a logical point of view, based on the noncontradiction principle, than by developing new positive doctrines. This places hylomorphism closer to the rising "analogical" mechanism of Descartes, Hobbes and Gassendi.

If we look for a good example of such a convergence, we may find it in the later work of the French *physicien* Marin Cureau de La Chambre (1594-1669). La Chambre was first doctor of Louis XIV of France, a friend of Descartes, Campanella and Fermat, and a founder of the French *Académie des sciences*. He is also the author of *Les Charactères des Passions* (1640-1662), one of the most influential treatises of physiognomy in the early modern age, as well as of an interesting book of psychology, *Le Systéme de l'âme*, published in 1664, in which he also concludes a famous controversy with Pierre Chanet on animal intelligence (Guidi 2015; Scribano 2016).

Starting from his first work, the *Nouvelles pensées,* published in 1634, Cureau tries to revitalize Marsilio Ficino's vision, founding it on a syncretic theory of spirits and light, and merging different elements from Renaissance pneumatology (especially Jean Fernel's), and of early modern mechanism. Like Ficino, and consistently with the long tradition of the "metaphysics of light" (Grosseteste), Cureau thinks of light as an intermediate between form and matter, and of spirit as a third substance between a material body and an intellective soul (Guidi 2016). However, La Chambre tries to justify his theory within the Jesuits' hylomorphic framework, as presented in some contemporary texts (Liceti 1616, 1641). According to Cureau, light can be defined as the act of a diaphanous body, which is its potential *substratum*. Here the form-matter relationship is directly matched to the act-potency one, making light a pure act-form, just as the diaphanous is a pure potency-matter of it. Thus, resorting to an (arguably) Dominican distinction between a *formal* and a *local* extension of bodies (Guidi 2017), La Chambre separates matter and extension, arguing that, even if they are nonmaterial bodies, light and spirits have a physical extension. He can thus establish that intermediate substances are direct physical manifestations of essences and forms, but, at the same time, he thinks of forms as the purest *efficient* causes in the physical world.

His theory of the intermediate substances is a turning point, from which Cureau can assure the ontological reality of catoptrics, proposing (especially in his latest works) an interesting mix of panpsychism and mechanism (Guidi 2016, forthcoming).

For Cureau, images are physical, extended, but non-material beings, made of reflected light rays, and they are, at the same time, pure extended essences in space. Thanks to images, Cureau argues, one can explain any *automatic* transfer of information in nature, without resorting either to Aristotelian μορφή or to Descartes' "extreme" reduction of form to "figure". La Chambre's theory of images, in fact, will meaningfully allow him to reintroduce a strong "activistic" view of imagination, proposing a kind of "catoptric mechanism". For him, imagination is a discursive, fully-automatic *ratio*, in which images are physical representations that can take the place of any scholastic *species intelligibilis*. They can also carry out any biological function previously ascribed to the vegetative or the sensitive soul.

Starting at least from his controversy with Chanet, Cureau uses images to "automatize" animals, affirming their intelligent – albeit preconditioned and instinctive – behaviour. Abiding by a Platonic innatism of sorts, Cureau puts forward a theory of mental images according to which animals have all their actions determined from birth by individual figures, impressed on their imaginations. In the instinctive behaviours, «les Images Naturelles» incline animals to their usual actions, including any sense-related action.

However, what is especially interesting in Cureau's theory of images is the function they play in his last work, *Le Système de l'âme*. Here La Chambre explicitly extends their activity to inanimate things, retracing Campanella's theory of *sensus*: also «dans les choses inanimées», «les Images font toutes seules» (Cureau de La Chambre 2004, p. 147). They are able to incline any object to automatically act and react, explaining several physical phenomena – for example magnetism (*ibidem*, pp. 135-152) – without using Aristotelian forms, as well as avoiding the reduction of nature to a *material* machine.

References

Ayers M. 1988, *Mechanism, Superaddition, and the Proof of God's Existence in Locke's Essay*, in «Philosophical Review», 90, pp. 210-251.
Boyle R. 1999-2000, *Works*, ed. by M. Hunter, E. B. Davis, 14 vols, Pickering & Chatto, London.
Carraud V. 2002, *Causa sive ratio. La raison de la cause, de Suarez à Leibniz*, PUF, Paris.
Coujou J.-P. 1999, *Suárez et la refondation de la métaphysique comme ontologie*, Éditions Peeters, Paris.
Courtine J.-F. 1990, *Suárez et le système de la métaphysique*, PUF, Paris.
Cureau de La Chambre M. 1634, *Nouvelles pensées sur les causes de la lumière, du desbordement du Nil et de l'amour d'inclination*, Pierre Rocolet, Paris.
Cureau de La Chambre M. 1640-62, *Les charactères des passions*, 5 voll., Iacques D'Allin, Paris.
Cureau de La Chambre M. 2004, *Le Système de l'ame*, Fayard, Paris.
Darmon A. 1985, *Les corps immatériels. Esprits et images dans l'oeuvre de Marin Cureau de La Chambre*, Vrin, Paris.
Des Chene D. 1996, *Physiologia. Natural Philosophy in Late Aristotelian and Cartesian Thought*, Cornell University Press, Ithaca & London.
Dijksterhuis E.J. 1961, *The Mechanization of the World Picture*, tr. by C. Dikshoorn, Oxford University Press, New York.
Esposito C. 2014, *Suárez and the Baroque Matrix of Modern Thought*, in *A Companion to Francisco Suárez*, ed. by V. Salas, R. Fastiggi, Brill, Leiden-Boston, pp. 124-147.
Forlivesi M., Boulnois O. 2002, *La distinction entre concept formel et concept objectif: Suárez, Pasqualigo, Mastri*, in «Les Etudes Philosophiques», 1(1), pp. 3-30.
Garber D. 2013, *Remarks on the Pre-history of the Mechanical Philosophy,* in Garber, Roux 2013, pp. 3-26.
Garber D., Roux S. (eds.) 2013, *The Mechanization of Natural Philosophy*, Springer, Dordrecht-.
G. Giglioni, G. 2013, *How Bacon Became Baconian*, in Garber, Roux 2013, pp. 27-54.
Guidi S. 2014, *La favola della materia. Epistemologia e narrazione nel* Monde *di Descartes*, in «Azimuth. Philosophical Coordinates in Modern and Contemporary Age», 4, pp. 83-113.
Guidi S. 2015, *L'angelo e la bestia. Metafisica dell'istinto tra Pierre Chanet e Marin Cureau de La Chambre*, in «Lo Sguardo», 18(2), pp. 233-258.

Guidi S. 2016, *Lo spirito e l'intelletto. Il Système de l'ame di Cureau de La Chambre*, in «Bruniana & Campanelliana», 22(1), pp. 215-225.
Guidi S. forthcoming, *L'estensione essenziale. La teoria della luce di Marin Cureau de La Chambre,* in «Physis».
Leijenhorst C.H. 2001, *The Mechanisation of Aristotelianism. The Late Aristotelian Setting of Thomas Hobbes Natural Philosophy*, Brill, Leiden.
Liceti, F., 1616, *De animarum coextensione corpori libri duo*, apud Petrum Bertellium, Padua.
Liceti F. 1641, *De luminis natura & efficientia, libri tres*, ex Typographia Nicolai Schiratti, Udine.
Rees G. 1975, *Francis Bacon's Semi-Paracelsian Cosmology*, in «Ambix», 22, pp. 82-101; 165-173.
Rees G. 1977, *Matter Theory: A Unifying factor in Bacon's Natural Philosophy?*, in «Ambix», 24, pp. 110-125.
Roux, S. 1996, *Les explications de la philosophie mécanique*, in «Lettre de la Maison Française d'Oxford», 4, pp. 125-136.
Sanhueza G. 1997, *La pensée biologique de Descartes dans ses rapports avec la philosophie scolastique. Le cas Gomez-Péreira*, L'Harmattan, Paris.
Scribano E. 2016, *Campanella's Coming Back. La Forge against Cureau de La Chambre*, in *Early Modern Philosophers and the Renaissance Legacy*, ed. by C. Muratori, G. Paganini, Springer, London.
Specht R. 1997, *Aspects 'cartésiens' de la théorie suarezienne de la matière*, in *Lire Descartes aujourd'hui*, ed. by O. Depré, D. Lories, Peeters, Louvain-la-Neuve, pp. 21-45.
Suárez, F. 1635, *Parti Secundae Summae Theologiae*, *Tractatus tertius de anima*, Lugduni.
Wadell M. 2015, *Jesuit Science and the End of Nature's Secrets*, Ashgate, USA.
Westfall R.S. 1977, *The Construction of Modern Science: Mechanisms and Mechanics*, Cambridge University Press, Cambridge.

Machines and Diseases: Giorgio Baglivi and his Mechanistic Physiopathology

Luca Tonetti

Introduction

The Croatian physician Giorgio Baglivi (1668-1707), professor of anatomy and surgery and, from 1702, of theoretical medicine at the Studium Urbis, played a pivotal role in the history of eighteenth-century Italian medicine (Salomon 1889; Scalzi 1889; Grmek 1960, 1991). His research was mainly concerned with the foundation of medical praxis on a Hippocratic and Baconian methodology (*De praxi medica*, 1696), and the definition of a fibrillary physiology able to explain morbid states too (*De fibra motrice et morbosa*, 1702).

As a supporter of the application of the laws of physics in medicine, and an opponent of ancient humoralism, Baglivi has been traditionally considered one of the leading representatives of the Italian iatromechanics and, in particular, of "fibre medicine" (Ishizuka 2012). This implies that fibres, as solids of the body, are the only parts responsible for the diseased condition. Morbidity is thus believed to be the result of an imbalance occurring in the oscillating fibres of the body.

Numerous 19th-century historians of medicine, such as De Renzi, Daremberg, Puccinotti, accepted this traditional interpretation of Baglivi's medicine. In support of this view, in *Machine et organisme* Georges Canguilhem quotes a famous passage from *De praxi medica*, in which Baglivi compares the parts of the body with different kinds of physical devices, such as retorts, hydraulic pipes, springs, ropes, and so on:

> Examine with some attention the physical economy of man: What do you find? The jaws armed with teeth: Are they anything but pliers? The stomach is but a retort; the veins, the arteries, the entire system of blood vessels are hydraulic tubes; the heart is a spring; the viscera are but filters, screens; the lungs are but bellows. And what are the muscles, if not cords? What is the ocular angle, if it is not a pulley? And so on. Let us leave it to chemists with their grand words of "fusion," of "sublimation," of "precipitation" to want to explain and thus to

> establish a separate philosophy; it is nonetheless incontestable that all these phenomena must be related to the laws of equilibrium, of angles, of cords, of the spring, and of the other elements of mechanics (quoted in Canguilhem 2009, p. 78).

According to Canguilhem, Baglivi proves that mechanistic explanations of bodily processes require engines, like springs, next to purely kinematic machines, which cannot properly describe the autonomy of organisms, as they only transform external energy. Baglivi seems to be a radical Cartesian physician, who overemphasises and extends Descartes' *bête-machine* argument to human beings.

However, contrary to what is generally claimed, Baglivi's physiology is not entirely compatible with a strict Cartesian view. Malpighi's emphasis on microstructures and the experiments on dura mater performed at Rome with Antonio Pacchioni (Conforti, De Renzi 2009) force Baglivi to "decentralize" Descartes' automaton. Baglivi introduces, in fact, not only different physiological processes (see, for example, the role of the nervous fluid and the distinction between *motus systalticus* and *motus reflexivus*), but also a different relationship between the heart and the meninges of the brain. Moreover, although he is an opponent of Paracelsus and Van Helmont, Baglivi does not exclude the role of chemistry, since for him life depends on a «complexum motuum chymico-mechanicorum» (Baglivi 1696, p. 97). Finally, while being a radical reductionist, Baglivi realizes that diseases cannot be completely reduced to mechanics, because they are something very different from machines malfunctioning.

Therefore, I believe that the traditional interpretation may support a stereotyped and naive image of Baglivi's work, and it needs a careful contextualization. In fact, in Baglivi the relationship between physics and medicine is much more multifaceted than is commonly acknowledged.

The aim of this short paper is twofold: firstly, I will examine the role played by mechanics in Baglivi's medicine in explaining normal as well as morbid states. I will then analyse the distinction between solids and fluids of the body. As an advocate of solidism, Baglivi focused both pathology and therapeutics mainly on the action of solids. However, fluids seem to still play an important role, because they are not easily reducible to the laws of mechanics and require rather specific approaches and methods.

Diseases and mechanical analogies

The "mechanistic turn" in pathology is well exemplified by Marcello Malpighi who, in his reply to Sbaraglia (Cavazza 1997), highlighted the need for a "multi-layered analysis" in medicine (Bertoloni Meli 2011), paying attention, in particular, to the role of mechanics. Malpighi suggested that mechanical devices or artefacts may allow physicians to reproduce morbid processes in order to understand – aprioristically – the underlying natural laws (Bertoloni Meli 2007, 2012, 2016). In fact, for Malpighi diseases are not a freak of nature, but a completely rational phenomenon (Bertoloni Meli 2001), whose properties can be deduced *a priori* from the knowledge of the structure and function of the affected bodily part as well as from the causes responsible for that disease.

Despite being a disciple of Malpighi, Baglivi follows a radically different approach. In fact, the relationship between physics (mechanics) and medicine is multifaceted. It is also worth noting that Baglivi never refers to mechanical devices as a means to describe and study morbid states. The only way to properly treat diseases is that of performing repeated and systematic bedside observations. Therefore, while supporting the importance of aetiology in medicine, Baglivi rejects the *a priori* method of rational medicine.

As already mentioned, one of Baglivi's main concerns was the crisis of medical practice, to which he devoted his first treatise, *De praxi medica*, published in 1696 with the aim of reforming medicine according to a Hippocratic and Baconian methodology. By sharing the same research method developed by Francis Bacon in *Novum Organum*, Baglivi listed six *idola* of medicine, namely, those impediments that have prevented progress in medicine. One of them, the third, concerns the use of analogies in medicine.

In order not to draw false conclusions, analogies should «relate only to things that fall under one Genus, as to Plants and Plants, Minerals and Minerals, Animals and Animals, & c. so that all the several Attributes of one thing may be verified of the other to which it is compar'd» (Baglivi 1704, p. 33). This rule allows physicians to underpin the extension of new findings in comparative anatomy across different species, because analogies are built «sub genere Viventium» (Baglivi 1696, p. 28). Moreover, thanks to the principle of uniformity, which states that the course of nature continues uniformly the same, analogies between machines and organisms can also be accepted. In fact, what the mechanics of solids and fluids suggests about the animate bodily structures is valid precisely because structures and functions of the body depend on number, weight and size, meaning that they respond to physical-mechanical properties. This is the reason why the analogies used by Giovanni Alfonso Borelli, Lorenzo Bellini (and Luca Tozzi) are supposed to be valid and useful for clinical practice (*ibidem*). Conversely, Helmontian, that is, chemical physicians do not build good analogies, because their comparisons are «extra sphaeram mutui praedicati» (*ibidem*, p. 29).

Baglivi himself used mechanical analogies: for example, in the *Epistola ad Alexandrum Pascoli* (1700), he argues that what allows muscles to contract is the *tomentum sanguineum*, i.e. the blood flowing between the fibrils. Baglivi assumes that blood corpuscles act like many fixed pulleys (or *trochleae*), around which the fibrils run as if they were strings (or levers). So, muscles can be regarded as a combination of simple machines, i.e. as a compound machine, whose mechanical advantage is the product of the mechanical advantages of the simple machines of which it is composed. In this case, given that the number of blood corpuscles is almost uncountable, the force amplification due to the *tomentum sanguineum* increases exponentially (Baglivi 1704, p. 402).

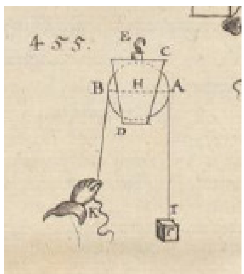


Figure 1. Diagram of a fixed pulley. Source: Schott 1661.

However, this hypothesis is untenable, because the fixed pulley does not properly represent the motility of blood corpuscles: therefore, Baglivi proposes a second image, by comparing corpuscles with a part of the wheel and axle, known as *scytala*, which is a sort of radial handle or lever that turns the wheel round.

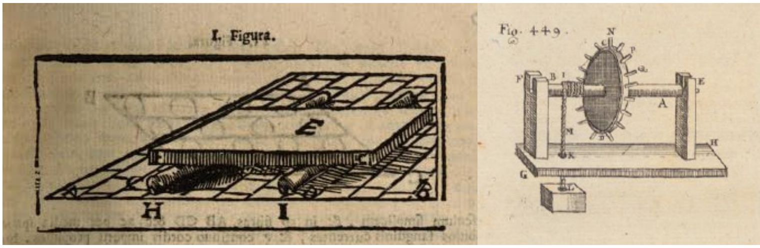


Figure 2. *To the left*, Baglivi's diagram of fibre, according to the second hypothesis. C and D are *scytalae*. Source: Baglivi 1700. *To the right*, diagram of wheel and axle. Source: Schott 1661.

In both cases, the mechanical analogy serves as a heuristic tool, through which Baglivi can identify the mechanism underlying muscular contraction and provide a new mechanistic explanation alternative to the prevailing one by Descartes.

Therefore, in order to make good assumptions, physicians should start with mechanical principles. However, unlike Malpighi, Baglivi does not rely on analogical reasoning in order to construct a natural knowledge *a priori.* Assumptions can promote clinical reasoning but, as they are simple theoretical constructions, they must be rejected when in conflict with experience.

Balancing fluids and solids of the body

Undoubtedly, mechanical analogies help understanding human physiology. But does this also apply to pathology? The problem of balancing fluids and solids of the body seems to show some limits in the mechanistic view on pathology.

In *De praxi medica* Baglivi recovers the traditional distinction between solids and fluids, without supporting the priority of solids over the rest of the body as he does in his later works on fibres. If all the diseases depended on a defect in solids, then physicians could easily find the causes and derive the necessary remedies, by applying the laws of physics. On the contrary, Baglivi believes that most of the diseases depend on fluids, whose constituents and internal structure are not easily reducible to mechanical patterns.

The difference between solids and fluids also influences the treatment of acute and chronic diseases. Acute diseases mainly affect the body fluids and, if not properly treated, they can easily become deadly or incurable. Physicians should limit the ministration of remedies and rely on the healing action of nature alone. Chronic diseases, however, may depend on the habit of solids or on those fluids delaying or failing to complete concoction. In this case, physicians are required to interfere with the natural course of disease, because nature, without the action of remedies, may not be able to resolve the morbid state.

In *De fibra* Baglivi outlines a more accomplished fibrillary theory, by distinguishing between two types of fibres, membranous and motor/muscular, which belong to two distinct but strictly interrelated subsystems, directed respectively by the *dura mater* (*dura mater*/nervous fluid/membranous fibres) and the heart (heart/blood/muscular fibres). Anyway, the identification of a microstructure (the fibre) responsible for diseases does not ensure the transition from physiology to pathology, and, still further, to therapeutics, which results from the attempt to combine fibrillary theory and Hippocratism.

This attempt is the main issue of Baglivi's commentary to Santorio's *De statica medicina*. Here Baglivi reinterprets the Hippocratic concept of balance as *eukrasia* in terms of an "equilibrium" of forces interacting between solids and fluids of the body. Fluids, although depending on the laws of solid mechanics, show their own properties and processes, such as digestion, concoction, fermentation, and so on. Only those who can properly assess the balance between both solids and fluids are able to treat diseases (can. XI: «Qui bene noverit aequilibrium inter solida oscillantia, & liquida currentia, morbos quamplures recte curare noverit»).

However, this balance is not the exact equilibrium of mechanics or hydraulics, but a "proportion" of sorts between solids and solids, fluids and fluids, solids and fluids, which cannot be deduced from the operations of mechanical devices, but only derived from clinical observations at the bedside and experiments. This is the reason why Baglivi prefers to test *in vivo* – for instance, by means of infusory surgery – because this enables him to evaluate solids and fluids simultaneously. In fact, thanks to the intravenous infusions, the remedy acts directly on both solids (vessel wall → membranous fibres) and fluids (blood).

Baglivi is not able to follow the rigor of mechanics or the axiomatic method of geometry in order to explain diseases. This does not mean that the physical and mathematical principles should be rejected or that they do not *inform* his medicine, but that, in some sense, pathology exceeds them: diseases are not simply malfunctioning machines. Since physicians are as yet unable to consider and quantify all the physical variables involved in a disease, such as the diameter of vessels, the force exerted by the heart, or the resistance of fluids and solids, machines cannot be useful as a direct reference model for pathology. As Baglivi says, the art of medicine can be improved only by use and exercise.

References

Baglivi G. 1696, *De praxi medica ad priscam observandi rationem revocanda*, *libri duo. Accedunt Dissertationes novae*, typis Dominici Antonii Herculis, sumptibus Caesaretti, Romae.

Baglivi G. 1700, *De fibra motrice, et morbosa; nec non de experimentis, ac morbis salivae, bilis, et sanguinis… Epistola ad Alexandrum Pascoli*, apud Costantinum, Perusiae.

Baglivi G. 1704, *The Practice of Physick, Reduc'd to the Ancient Way of Observation*, Andrew Bell, Ralph Smith, and Daniel Midwinter, London.

Baglivi G. 1704, *Opera omnia medico-practica, et anatomica*, sumptibus Anisson, & Joannis Posuel, Lugduni.

Bertoloni Meli D. 2001, *Blood, Monsters, and Necessity in Malpighi's De Polypo Cordis*, in «Medical History», 45(4), pp. 511-522.

Bertoloni Meli D. 2006, *Thinking with Objects. The Transformation of Mechanics in the Seventeenth Century*, The Johns Hopkins University Press, Baltimore.

Bertoloni Meli D. 2007, *Mechanistic Pathology and Therapy in the Medical* Assayer *of Marcello Malpighi*, in «Medical History», 51, pp. 165-180.

Bertoloni Meli D. 2011, *Mechanism, Experiment, Disease. Marcello Malpighi and Seventeenth-Century Anatomy*, Johns Hopkins University Press, Baltimore.

Bertoloni Meli D. 2012, *Machines and the Body between Anatomy and Pathology*, in *Modèle Métaphore Machine Merveille*, ed. by A. Gaillard, J.-Y. Goffi, B. Roukhomovsky, S. Roux, Presses Universitaires de Bordeaux, Bordeaux, pp. 53-68.

Bertoloni Meli D. 2016, *Machines of the Body in the Seventeenth Century*, in *Early Modern Medicine and Natural Philosophy*, ed. by P. Distelzweig, B. Goldberg, E.R. Ragland, Springer, Dordrecht-London, pp. 91-116.

Canguilhem G. 2009, *Knowledge of Life*, transl. by S. Geroulanos, D. Ginsburg, Fordham University Press, New York.

Cavazza M. 1997, *The Uselessness of Anatomy: Mini and Sbaraglia versus Malpighi*, in *Malpighi Anatomist and Physician*, ed. by D. Bertoloni Meli, Olschki, Firenze, pp. 129-145.

Conforti M., De Renzi S. 2009, *Sapere anatomico negli ospedali romani: Formazione dei chirurghi e pratiche sperimentali (1620-1720)*, in *Rome et la science moderne: Entre Renaissance et Lumières*, ed. by A. Romano, Publications de l'École française de Rome, Rome, pp. 433-472.

Dell'Anna G. 1990, *Giorgio Baglivi e la «Medendi methodus»*: una rilettura dell'empirismo baconiano, in *Medicina e biologia nella rivoluzione scientifica*, ed. by L. Conti, Edizioni Porziuncola, Santa Maria degli Angeli-Assisi, pp. 272-288.

French R.K. 2003, *Medicine Before Science: The Business of Medicine from the Middle Ages to the Enlightenment*, Cambridge University Press, Cambridge.

Grmek M.D. 1960, *Osservazioni sulla vita, opera ed importanza storica di Giorgio Baglivi*, in *Atti del 14. Congresso internazionale di storia della medicina*, Roma-Salerno, 13-20 settembre 1954, Guerra e Belli, Roma, pp. 423-437.

Grmek M.D. 1991, *La vita e l'opera di Giorgio Baglivi medico raguseo e leccese (1668-1707)*, in *Il nucleo filosofico della scienza*, ed. by G. Cimino, U. Sanzo, G. Sava, Congedo, Galatina, pp. 93-111.

Ishizuka H. 2012, *"Fibre Body": The Concept of Fibre in Eighteenth-century Medicine, c.1700–40*, in «Medical History», 56(4), pp. 562–584.

Müller I.W. 1991, *Der Hippokratismus des Giorgio Baglivi*, in «Medizinhistorisches Journal», 26(3-4), pp. 300-314.

Salomon M. 1889, *Giorgio Baglivi und seine Zeit. Ein Beitrag zur Geschichte der Medicin im 17. Jahrhundert,* Hirschwald, Berlin.

Scalzi F. 1889, *Giorgio Baglivi e il suo tempo*, in «Lo Spallanzani», VII-VIII, pp. 321-337.

Schott G. 1661, *Cursus mathematicus, sive absoluta omnium mathematicarum disciplinarum Encyclopaedia, in libros XXVIII digesta*, sumptibus haeredum Joannis Godefridi Schönwetteri bibliopolae Francofurtensis. Excudebat Jobus Hertz typographus Herbipolensis, Herbipoli.

Tonetti L. 2016, *La sfida della causalità alla pratica medica: il modello eziologico galenico e il dibattito medico tardo-seicentesco*, in *Percorsi evolutivi. Lezioni di Filosofia della biologia*, ed. by E. Gagliasso, F. Morganti, A. Passariello, FrancoAngeli, Milano, pp. 131-143.

Toscano A. 2004, *Mirabilis machina: il perpetuum mobile attraverso il De statice æris e il De fibra motrice et morbosa di Giorgio Baglivi*, Brenner, Cosenza.

Vidal M. 1990, *Giorgio Baglivi tra osservazione clinica e speculazioni iatromeccaniche*, in «Atti del centro ricerche storiche di Rovigno», 20, pp. 133-214.

Early Organicism and its Juggling Machines: Further from Nature, Closer to Organisms

Alessandra Passariello

Introduction

In 20th century developmental studies, the debate on the epistemological equivalence/difference between machines and organisms has a tangled genealogy. The debate first burst onto the scene in the late-1920s with the rise of the Organicist movement, which supported the idea that living beings are radically different from machines (Woodger 1929; von Bertalanffy 1933, 1952).

However, I argue that the Organicist standpoint in the machine-organism debate can only be grasped if we extend our analysis to the broader domain of relations between organisms, machines and natural phenomena.

Early Organicism claims a qualitative distinction between biological and artificial phenomena in the same way it calls for a clear-cut distinction between biological and natural (physical-chemical) ones. At first glance, this would only appear to separate organisms from the realms of both natural and artificial phenomena but, on a deeper look, it implies an epistemological similarity between the latter two.

In his earlier texts (1933, 1952), von Bertalanffy assumed an epistemological proximity between machines and natural systems utilising an older concept of machine as «an arrangement of processes» determined by «a fixed structure», thus a machine as a physically reducible entity (von Bertalanffy 1952). Consequently, he drew the distinction between organisms and machines with the intention of avoiding the physical-chemical reducibility of organisms. However, at the time von Bertalanffy was writing, a major shift in the relation between machines and natural systems was about to occur.

From the 1940s on, the divide between physical-chemical and artificial systems brought to light by Information Theory, Cybernetics and Computer

Science, would bring machines, in their renewed form of organized, hierarchical dual control systems, closer to organisms (Polanyi 1968).

Von Bertalanffy's and the machine theoretical conception of development

In the 1920s and 1930s, a geographically distributed network of scientists used a common label for their work: "organicism" which meant a different biological interpretation regarding

- a new experimental methodology, privileging an enquiry onto the entire organism and not on its disassembled parts;
- a new type of causal explanation, between a physical-chemical and a supposed metaphysical, vitalistic one.

The rise of the organism-machine debate may be framed within the latter epistemological concern for a causal explanation that differs from both physical-chemical reductionism and metaphysical vitalism. This antithesis was one between a material causality, rigorously bound to the explanatory domain of physical-chemical laws, and an immaterial causality, calling for literally "metaphysical" (beyond physical laws) explanatory principles. One may notice that in this epistemological debate that sets those we may call the "mechanists" (and later "neo-mechanists") against the much-derided vitalists (and "neo-vitalists"), the relation between organisms and machines is not the main controversial point. Indeed, the main issue is the epistemological proximity/distance between natural (physical-chemical) and biological phenomena. So, when do machines make their entrance?

In his *Problems of Life* (1952) von Bertalanffy critically points out at three "leading ideas" of what was at the time modern biological thought: they are the *analytical-summative* claim, the *machine-theoretical* and the *reaction-theoretical* ones[1] (von Bertalanffy 1952, p. 9). According to von Bertalanffy, mechanists adopt an *analytical-summative* standpoint: they analyse biological phenomena (e.g. metabolism, development) by decomposing them into different parts and by enquiring separately into the properties of those parts in isolation from each other. Once the properties exhibited by the parts in isolation have been explained, the overall biological phenomenon is supposed to

[1] According to von Bertalanffy, the *reaction-theoretical* assumption considers the organism as an "automaton", that is, «a passive system, set into action through outside influences, the so-called stimuli» (von Bertalanffy 1952, p. 18). This assumption deals with the then current explanation of animal behaviour and will not be addressed in the article.

result from the aggregation of their properties in what is labelled as a summative way. The analytical, decomposing stance and the summative, aggregative one are thus two complementary faces of the same method.

Von Bertalanffy harshly criticizes this methodological protocol both by reporting some counterexamples and by giving more general theoretical reasons. In the first instance, a telling counterexample is the case of tissue culture: «If cells are explanted from the organism and allowed to grow as a tissue culture in an appropriate nutrient, their behaviour will be different from that within the organism» (*ibidem,* p. 12). From a theoretical point of view, von Bertalanffy suggests that most primary causal factors have to be searched for in the relations among the organism's parts and not in their individual properties.

It is worth noting that the rejection of the analytical-summative method is not a rejection of the validity of the physical-chemical laws (regularities) in biological phenomena. However, according to von Bertalanffy, since those laws can only be used as analytical tools, i.e. in explaining the properties of isolated organisms' parts as if they were ordinary physical-chemical phenomena, they are insufficient and, most of all, misleading for the search for an explanation of biological phenomena. Von Bertalanffy argues that only the formulation of exact laws concerning the whole system i.e. the set of interactions between its components parts, can properly explain biological properties. But, in his view, this approach doesn't yet put the machine-analogy explicitly under indictment.

Interestingly, in the same text von Bertalanffy couples his criticism of the *analytical-summative* method to another argument against the *machine-theoretical* concept. This second concept, he argues, is shared by both mechanists and vitalists and is based on the idea that «order in vital phenomena» can be «interpreted in terms of structures», that is, physical-chemical «mechanisms in the widest sense» (*ibidem*, p. 16). The emphasis on a machine's dependence on its structure may sound misleading. We could be tempted to see in the term "structure", here referring to machines, an analogue of the concept of higher-level boundary conditions (Polanyi 1968) harnessing physical-chemical processes. This interpretation of machines as what later became known as "dual control hierarchies" would be entirely anachronistic with respect to von Bertalanffy's text.

Arguably, until the rise of cybernetics (Rosenblueth, Wiener, Bigelow 1943) scientists conceived of machines as a strict application of physical-chemical laws: in other words, in a machine «the arrangement of processes can be explained […] only by assuming that the processes are directed in the right way by a fixed structure» (*ibidem*, p. 6). According to von Bertalanffy, machine functioning is so highly dependent on their physical-chemical structure that any alteration of it prevents the machine from achieving its task or, adopting a less teleological language, from showing its characteristic properties.

As for the analytical-summative stance, in this case too von Bertalanffy brings some counterexamples and clarifies them through more general, theoretical arguments. He resorts directly to Driesch's well-known experiment on the isolated sea urchin's blastomeres and shows how at the end of their biological development a fully developed sea urchin larva is achieved, despite many external disturbances altering structure during its embryonic development.

It is worth taking a brief look into Driesch's experiments on regulative embryos. The experiment was performed on fertilized eggs of a sea urchin species known as *Echinus tuberculatus* and consisted in separating the first two blastomers (cells resulting from the first zygote's mitosis) from each other. Basing his expectations on Roux's experimentally based concept of "mosaic development" (Figure 1) and on Weismann's explanation of development through the mechanism of "(qualitative) unequal nuclear divisions" (Figure 2), Driesch expected to observe the production of two half-embryos. However, to his surprise, at the gastrula stage, development «yielded a complete individual of half-size (dwarf-gastrula)» (quoted in Sander 1997a, p. 178).

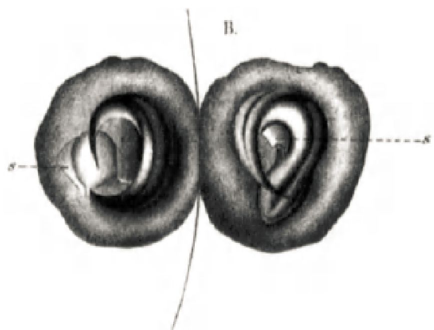


Figure 1. Mosaic development. Source: Roux 1888.

Figure 2. Qualitative unequal nuclear divisions. Source: Weismann 1892.

This showed that the eggs of *Echinus tuberculatus* could regulate their development: to Driesch, this meant that cell differentiation is achieved through reciprocal interactions, not through a process of mosaic, independent development. If, as Roux had hypothesized, development was a matter of “independent differentiation” and every cell was only provided with the causal factors accounting for its future morphological and functional characters, those parts of the embryo deriving from the one blastomere could not be obtained from the other. Regulation – that is, the occurrence of a causal interaction between cells – is not a problem *per se* but it becomes a major issue with respect to the validity of the machine analogy; this is particularly true when the “cellular interactions”, which should be firmly anchored to their physical-chemical structure, are robust enough to persist despite a radical change in the same physical-chemical structure.

In fact, Driesch's experiment shows not only that development needs cellular interactions and is thus regulative rather than mosaic-like; it also, most importantly, shows that the organism is capable of rebuilding interactions even when an experimentally induced disturbance radically modifies its physical-chemical structure. It is this kind of regulation, intended as a robustness to physical-chemical alterations, which, according to von Bertalanffy, makes the machine-analogy unsuitable to describe and interpret development. Both Driesch and von Bertalanffy refer to this robustness through the word "regulation" and both of them consider such a regulative capacity not ascribable to machines.

In this perspective, it seems to us that, according to von Bertalanffy, machines are reducible to physical-chemical laws precisely *because* they are strictly dependent on their physical-chemical structure. Unlike machines, organisms can cope with alterations in their physical-chemical structure and arguably possess this kind of regulative capacity because of systems regularities which cannot be investigated through the *analytical-summative* method.

However, von Bertalanffy seems to be aware that regulation, ever though a characteristic property of biological phenomena, comes in various degrees «ontogenetically as well as phylogenetically» (von Bertalanffy 1952, p. 17). Given the above-mentioned regulative properties, «organisms are not machines»; but they «can to a certain extent become machines, congeal into machines» when the relation between their distinctive properties and their structure becomes one of a bi-univocal nature. In those cases, fully illustrated by the progressive decrease in regulative capacities during ontogeny, minimal alterations in the organism's structure will result in a change in its properties.

Despite this quantitative account of regulation and the implicit possibility of ranking organisms according to a measure of their robustness, the relation between organisms and machines remains the one of a qualitative distinction. Though organisms face a «transition from less mechanized and more regulative states to more mechanized and less regulative ones», this is never fully accomplished. The fact that an organism faces a «perpetual breaking down and replacement of its building materials» could be interpreted in von Bertalanffy's view as the minimum degree of regulation an organism can achieve; of course, the measure a machine (according to the definition of von Bertalanffy) would achieve is > 0.

In other words, as long as machines are investigated from the standpoint of their structure, i.e. as a set of physical-chemical ordered events, regulation, defined in terms of the constancy of properties in the presence of a change in structure, cannot be ascribed to them.

Conclusions

As anticipated in the introduction, a major shift in the concept of the machine was about to occur at the time von Bertalanffy was writing. Information Theory, Cybernetics and Computer Science proposed a new way of looking at machines as organized systems with relationships to the physical-chemical substrate that were either irrelevant or at the very least interchangeable.

In particular, the advances in Information Theory and Computer Science gave birth to a new definition of "organization" in terms of transmission and elaboration of a certain amount of information while cybernetics pointed to a mechanism, the feedback loop, able to reintegrate, in terms of information, the end-result (the elaborated information) of the machine.

In a sense, for machines at least, von Bertalanffy's concept of "organization" as a property that cannot be reduced to physical-chemical interactions has been formulated in terms of the concepts of transmission, elaboration and control of a finite amount of information. Of course, this shift in the relation between machines and natural phenomena does not necessarily imply reducing von Bertalanffy's notion of biological organization to the informational/cybernetic/computational organization of machines.

However, I suggest that organicist arguments, eventually rejecting the epistemological analogy between machines and organisms, must take the informational/cybernetic/computational shift seriously. This means that taking an organicist standpoint today, after the development of cybernetics, implies acknowledging that despite eventual differences in the biological and the machine notions of organization, organisms and machines are both organized systems with properties that cannot be explained through the application of purely physical-chemical laws. This also means acknowledging that machines have successfully jostled in the interspace between the natural and the biological, now being further from thunderstorms and closer to organisms (Keller 2008).

References

Bertalanffy L. von 1933, *Modern Theories of Development: An Introduction to Theoretical Biology*, transl. by J.H. Woodger, Oxford University Press, Oxford.

Bertalanffy L. von 1952, *Problems of Life: An Evaluation of Modern Biological and Scientific Thought*, Harper & Brothers, New York.

Keller E.F. 2008, *Organisms, Machines and Thunderstorms: a History of Self-Organization. Part One*, in «Historical Studies in the Natural Sciences», 38(1), pp. 45-75.

Polanyi M. 1968, *Life's Irreducible Structure*, in «*Science*», 160, pp. 1308-1312.

Rosenblueth A., Wiener N., Bigelow J. 1943, *Behaviour, Purpose and Teleology*, in «Philosophy of Science», 10, pp. 18-24.

Roux W. 1888, *Beiträge zur Entwickelungsmechanik des Embryo: V. Über die künstliche Hervorbringung halber Embryonen durch die Zerstörung einer der beiden ersten Furchungszellen, sowie über die Nachentwicklung (Postgeneration) der fehlenden Korperhälfte*, in «Virchows Archiv», 114, pp. 113-153, 246, 291 (engl. transl. in *Foundations of Experimental Embryology*, ed. by B.H. Willier, M.J. Oppenheimer, Englewoods Cliffs, Prentice-Hall).

Roux W. 1895, *Einleitung*, in «Archiv für Entwicklungsmechanik der Organismen», 1, pp. 1-42 (engl. transl. by W.M. Wheeler, *The Problems, Methods and Scope of Developmental Mechanics*, in *Biological Lectures Delivered at the Marine Biological Laboratory of Wood's Hole*, Ginn, Boston 1894, pp. 149-190).

Sander K. 1997a, *Shaking a Concept: Hans Driesch and the varied Fates of sea Urchin Blastomeres*, in *Landmarks in Developmental Biology 1883-1924*, ed. by K. Sander, Springer-Verlag, Berlin, pp. 29-31.

Sander K. 1997b, *Hans Driesch and the Critical Mechanist: Analytische Theorie der organischen Entwicklung*, in *Landmarks in Developmental Biology 1883-1924*, ed. by K. Sander, Springer-Verlag, Berlin, pp. 30-32.

Sander K. 1997c, *Entelechy and the Ontogenetic Machine: Work and Views of Hans Driesch from 1895 to 1910*, in *Landmarks in Developmental Biology 1883-1924*, ed. by K. Sander, Springer-Verlag, Berlin, pp. 36-38.

Weismann A. 1893, *The Germ-Plasm: A Theory of Heredity*, transl. by W.N. Parker, H. Rönnfeldt, Scott, London.

Woodger J.H. 1929, *Biological Principles: A Critical Study*, Routledge & Kegan Paul, London.

From Brains as Machines to Machines as Brains. A Short Historical and Epistemological Reflection on the Simulation and "Reverse Engineering" of the Central Nervous System

Mattia Della Rocca

The always-varying relationship between science and technology has been a classical topic of interest to the 20th century historiography and philosophy of science – probably, in some ways, it has even been a foundational issue for both fields. As it is well known, some pivotal works on this subject put the dialectics between science and technology at the core of their analysis (Schuhl 1947; Koyré 1961; Rossi 1962), drawing the attention of many scholars in these fields upon this complex (and often controversial) subject.

However, this relationship has also become a core issue for the historical and epistemological inquiry on the 21st century life sciences and related technologies. Although not all fields in the vast panorama of life sciences are equally affected by this trend, one can observe a growing tendency within large sectors of present-days biomedicine to shift towards the epistemic and methodological assumption that «biology *is* technology», to quote the title of a recent book published by Harvard University Press – a volume significantly written not by a scholar from the academic world, but by a CEO of *Biodesic LLC,* a bioengineering firm based in Seattle (Carlson 2010). In more recent years, also neuroscience seems to have enthusiastically joined this "technological turn" in life sciences, arguing for the benefits of its technological translation in several application areas, such as industry, medicine and everyday-life. Furthermore, this technological turn seems to have deeply transformed neuroscience keywords and core concepts, in order to fit them into the vocabulary of the new technologies, especially those related to the Information and Communication Technology field (ICT). This "technological shift" seems to have gained a particular prominence with the advent of the "big brain science" projects, the *BRAIN Initiative* (BI) and the *Human Brain Project* (HBP), both launched in 2013, respectively in the United States and in the European Union.

Within these projects, technology represents more than a simple instrument for gaining new knowledge on the brain. Indeed, 21st century "big brain science" projects – and especially HBP – overtly aim to achieve a simulation of the central nervous system by means of "reverse engineering", which should enable the creation of a new class of hardware labelled as "neuromorphic" (that is, technology meant to have the "form of the brain"). Reverse engineering and neuromorphic computing are fundamental missions for the Human Brain Project. The development of neuromorphic computing hardware and software represents one amongst the six ICT platforms created within the project – i.e. the "Neuromorphic Computing Platform". The Neuromorphic Computing Platform is tightly linked with three other HBP platforms, the "Neuroinformatics Platform", the "Brain Simulation Platform" and the "High-Performance Computing Platform". The first should allow neuroscientists involved in the project to analyse and predict brain processes, while the second is conceived for implementing and testing brain models through neuromorphic computation; the third provides the computational capabilities and software needed «to create, simulate and analyse multiscale brain models» (Calimera *et al.* 2013, p. 193). Within this international cooperative project, reverse engineering of the brain and neuromorphic hardware are often presented as the "next revolution" in neuroscience, a «scientific prize» which will catalyse a «methodological paradigm shift» – in the words of one of HBP's leaders, Richard Frackowiak – that will change our understanding of both neuroscience as a scientific field and of its object of inquiry (i.e., the central nervous system) (Frackowiak 2014).

Lately, however, such an opinion has been repeatedly challenged by neuroscientists and by philosophers of science as well; the former criticizing the lack of theoretical reflection on the simulated models, the latter denouncing HBP project's optimism as wishful thinking while pointing out the risky and "ideological" features of such a theoretical and methodological approach (Frégnac, Laurent 2014; Haueis, Slaby 2015; Della Rocca 2015, 2017).

This said, what does it mean to simulate a brain and create a neuromorphic hardware? What does the adoption of an engineering perspective in neuroscience imply for the contemporary Western discourse about the brain/mind system? And further, are these technologically-oriented goals actually a new perspective in the field of brain research? In this short historical and epistemological review of the topic, I will try to provide an answer to these questions by exploring whether this "technoscientific" approach to the brain actually represents a novelty in the scientific panorama – and, if it does, in respect to what.

Let us turn to "neuromorphism" to provide a first answer to these questions. The term "neuromorphic" – a neologism to indicate a technological apparatus which imitates the structure and the function of the brain – has been coined by the US scientist and engineer Carver Mead (born 1934), at the beginning of the 1990s (Mead 1990). Contemporary research on neuromorphic devices conceptually stemmed from the development of Very-Large-Scale-Integration (VLSI) technologies, whose first concept was formulated around the early 1970s. VLSI research aimed to create integrated circuits by combining thousands of transistors into a single chip – as in the case of present-day microprocessors, which combine in one single chip different circuitry as for CPU, ROM or RAM – and emphasized the non-linear characteristics of the transistor. This area of research gained a concrete degree of feasibility by the pioneering studies, led around the mid-1980s by prominent scientists as John Hopfield, Carver Mead and Richard Feynman (Hey 1999). Inspired by the research on biological visual systems, Mead took interest in replicating the graded synaptic transmission of the retina, exploiting the analogue properties of transistors rather than using them in a classical digital way. A year before the creation of the neologism "neuromorphic hardware", Mead had showed how such an analog circuit shared many common physical properties with protein channels in neurons (Mead 1989), and that these types of circuits require far fewer transistors to emulate the functions of neural systems, compared with the digital ones. In time, neuromorphic chips have also implemented some mechanisms that can easily modify their electronic "synapses" as soon as data are processed, simulating the brain's plasticity (Indiveri, Horiuchi 2011).

Conventionally, the early history of the implementation of neural circuits in electronic models is traced back to the construction of perceptrons by Frank Rosenblatt (Rosenblatt 1958), followed by the development of artificial electronic retinas by Kunihiko Fukushima (Fukushima *et al.* 1970). However, from the perspective of their "material" ancestry neuromorphic devices originated elsewhere. In fact, they stemmed from the development of physical models of the nerve – as Ralph Lillie's iron-wire model of the 1920s, which simulated the biophysics of action potentials in a synthetic system. Furthermore, epistemologically speaking, they draw on the identity postulated between biological and electronic components (as materialized in the 1940s by Kenneth Cole in his "voltage clamp" technique), as well as from the functional equivalence between living beings and machines in the "biomimetics" of the early 1950s (as in the works of Otto Herbert Schmitt). Neuromorphism, indeed, has a longer history, which predates and flows parallel to that of "connectionist" or "post-classical" Artificial Intelligence.

In a similar way, the mid-1980s also saw the official birth of the "reverse engineering" concept, which was tightly connected to the development of "biologically realistic machines" since its beginnings. The first occurrence of the term in overt relation with the human brain can be found in a 1985 article by John King McIanahan Stevens – significantly appeared in the pages of *BYTE Magazine*, an American magazine dedicated to microcomputing – in which the author asked his reader how to «develop more efficient "sixth-generation" artificial-intelligence (AI) computers using circuitry copied directly from the brain, perhaps using radical new architecture, new hardware, and an entirely new logic» (Stevens 1985, p. 287).

However, the very concept of "reverse engineering" was adopted earlier by scholars from mixed backgrounds, such as Edwin Lewis, who in the early 1960s worked with Richard Reiss and Theodore Bullock on the development of a nerve net simulation, which was aimed to achieve a high degree of realism in the electronic simulation of a part of the central nervous systems. Lewis' work and ideas fitted completely in what can be seen as the first phase of the "technoscientific" approach to the brain, in which researchers from several disciplines – mostly biology, neurophysiology and engineering – tried to achieve, with the development of this type of analogies, a twofold goal: on one hand, to develop artificial systems which duplicate as closely as possible the physiological events which take place *in vivo* in order to understand more about the underlying physiological or chemical events, and, on the other one, to use simulation of nerves in studying the manipulation of information for the subsequent development of new machines (i.e. computers) (Harmon 1961). It is worth noting that this "technoscientific approach to the brain" experienced its own epistemological crisis – i.e. in the 1970s, when the very emphasis on the axonal level became an obstacle to the development of realistic simulation. This "crisis" led to a stop in the development of neuromorphic technologies until its resumption by Mead and colleagues in the late 1980s but, at the same time, it bolstered a deeper reflection on the role and the limits of brain models, eventually leading to the "plasticity revolution" of the late 1980s (Della Rocca *in press*; Morabito *in press*).

Rather than a new paradigm in contemporary neuroscience, thus, reverse engineering and neuromorphism should be more correctly considered as one of the epistemological and methodological pillars of the field since its very beginnings. Therefore, it is difficult to hail them as symptoms of the breakthrough of a "methodological paradigm shift", as promoted and propagandized by 21st century "big brain science" projects' rhetoric.

In the light of the historical and epistemological inquiry, neuromorphism and reverse engineering of the central nervous system appear to be part of a long tradition in contemporary neuroscience, and they rather urge our "brain society" to reflect on the epistemological and cultural transformation which our idea of the brain has undergone in the last decades. In particular, I am referring to the shift from the "brain as a machine" analogy (proper of cybernetics and scientific popular culture of the second half of the 20th century) to the "machine as a brain" one (which today dominates our epistemic landscape and technological imagery). Clearly, this epistemic inversion affects our scientific categories in neuroscience (especially about modelling), as well as the expectations on brain research *tout court*. Thus, a reflection on this issue cannot be avoided, both for its societal-cultural implications and for the relevance it could have on the epistemological debate on neuroscience *per se*. Indeed, the real "prize" for this reflection – to use Frackowiak's word – is a critical acknowledgement of the very advantages and limitations of the technoscientific approach to the brain: for what is at stake is the development of a historical and philosophical awareness, a necessary step if the science-culture-society dialectic wants to avoid another "crisis in (present-day) neuroscience".

References

Calimera A., Macii E., Poncino M. 2013, *The Human Brain Project and Neuromorphic Computing*, in «Functional Neurology», 28(3), pp. 191-196.

Carlson R.H. 2010, *Biology Is Technology. The Promise, Peril, and New Business of Engineering Life*, Harvard University Press, Cambridge, MA.

Cole K.S. 1949, *Some Physical Aspects of Bioelectric Phenomena*, in «Proceedings of the National Academy of Sciences», 35(10), pp. 558-566.

Della Rocca M. 2015, *Ideologia delle/nelle neuroscienze contemporanee,* in *Per una scienza critica*, ed. by E. Gagliasso, M. Della Rocca, R. Memoli, ETS, Pisa, pp. 189-199.

Della Rocca M. 2016, *Sulla questione ideologica nelle neuroscienze,* in *Percorsi Evolutivi. Lezioni di Filosofia della biologia*, ed. by E. Gagliasso, F. Morganti, A. Passariello, FrancoAngeli, Milano, pp. 189-199.

Della Rocca M. *in press*, *Histories of the Brain. Towards a Critical Interaction of the Humanities and the Neurosciences*, in *The Human Sciences after the Decade of the Brain*, ed. by J. Leefmann, E. Hildt, Academic Press, Cambridge, MA.
Frackowiak R. 2014, *Eyes on the Prize*, in «New Scientist», 223(2978), pp. 28-29.
Frégnac Y., Laurent G. 2014, *Where is the Brain in the Human Brain Project?*, in «Nature», 513(7516), pp. 27-29.
Fukushima K., Yamaguchi Y., Yasuda M., Nagata S. 1970, *An Electronic Model of the Retina*, in «Proceedings of the IEEE», 58, pp. 1950-1951.
Harmon L.D. 1961, *Studies with Artificial Neurons, I: Properties and Functions of an Artificial Neuron*, in «Kybernetik», 1(3), pp. 89-101.
Haueis P., Slaby J. 2015, *Brain in the Shell: Assessing the Stakes and the Transformative Potential of the Human Brain* P*roject*, in *Neuroscience and Critique,* ed. by J. De Vos, E. Pluth, Routledge, London, pp. 117-140.
Hey T. 1999, *Richard Feynman and Computation*, in «Contemporary Physics», 40, pp. 257-265.
Indiveri G., Horiuchi T.K. 2011, *Frontiers in Neuromorphic Engineering*, in «Frontiers in Neuroscience», 5:118.
Lillie R.S. 1922, *Transmission of Physiological Influence in Protoplasmic Systems, especially Nerve*, in «Physiological Reviews», 2(1), pp. 1-37.
Mead C.A. 1989, *Analog VLSI and Neural Systems*, Addison–Wesley, Reading.
Mead C.A. 1990, *Neuromorphic Electronic Systems*, in «Proceedings of the IEEE», 78(10), pp. 1629-1636.
Morabito C. *in press*, *Plastic Maps: the New Brain Cartographies of 21th Century Neurosciences*, in «Nuncius. Journal of the Material and Visual History of Science», 32(3).
Rosenblatt F. 1958, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, in «Psychological review», 65(6), pp. 386-408.
Stevens J.K. 1985, *Reverse Engineering the Brain*, in «BYTE Magazine», 10(4), pp. 287-299.

II

Understanding by Building. The Use of Simulation in Cognitive Science

Understand Me or Duplicate Me? Levels of Explanation in Artificial Methodology

Nicole Dalia Cilia

The development of new technologies has caused a proliferation of use of "machines" in different venues and fields, such as households, medicine, and scientific research. A well-programmed machine is able, if not to replace humans, to assist them in many tasks and to replicate, and sometimes explain, many cognitive processes. Nowadays, however, research is trying to reach a more ambitious bigger result. The study of intelligent behaviour so as to uncover hidden mechanisms is not any more a goal *per se*; what is at stake is the artificial reproduction of the entire human brain. Contemporary neuroscience has seen major research projects aiming to achieve large-scale *simulations* of brain mechanisms, characterized by previously unattainable levels of accuracy.

As underlined by Santucci *et al.* (2016), the key idea of the "simulative method" in the artificial sciences (Simon 1996) is «understanding by building». If we consider the simulative approach at large, including the simulation of physical, economic or social phenomena, this method is now widely in use, not only in the natural sciences, but in many other domains as well. If we focus, instead, on the more specific sense in which the simulative approach was originally intended in cognitive science and AI – that is, on the idea of building or reproducing biological intelligence in the form of (adaptive) machines – the status of this method is far from being clear or consolidated in the scientific community (Arbib 2003; Oudeyer 2010). In order to analyze the different methodologies used in the cognitive sciences, they will be grouped into two broad families, reflecting two different research goals: *model-oriented* and *data-oriented* simulations.

Model- and *data-oriented* simulation

Since the 1940s, in order to reproduce and investigate the mechanisms of cognitive functions, cognitive scientists have followed a methodology known as *synthetic method* (Cordeschi 2002). Following this approach, machines are used for testing theories, since a "mechanical organism" (or machine) and a "biological organism" (or proper body) seem to share some «essential characteristics of the investigated phenomenon, revealing a common functional organization beyond the different physical structures» (Cordeschi 2008, *passim,* my translation). Therefore, the goal of the synthetic method is testing the "mechanisms" by which machines are built, not reproducing cognitive functions. This is made possible by the comparison between machine and human behaviours. According to Cordeschi (2000), the first explicit attempt to apply the synthetic method was the machine described by S.B. Russell in 1913, thirty years before the publication of Rosenblueth, Wiener and Bigelow (1943). This machine was a hydraulic device which simulated a few forms of associative learning. The modeling methodology included the three steps that characterize today the synthetic method: 1) the formulation of hypotheses; 2) the description of the machine that "incorporated" that assumptions; 3) the comparison of the results obtained from the machine with those of the organism. This machine was «designed so as to embody certain hypotheses on the plasticity of nervous connections pointed out at the time by psychologists in order to explain the physical bases of learning» (Cordeschi 2000, p. 315).

This was in itself a major shift, because a machine able to modify its behaviour in relation to the environment – that is, able to learn – demanded at the time an expansion of the machine concept such that it could represent a«test […] of a theory, because machine and organism shared some essential features of the characteristic investigated» (Cordeschi 2008, p. 170, my translation).

Another author who stressed the importance of the synthetic method was Kenneth Craik. In 1943 he claimed that there is a difference between the "analytic method", which investigates the anatomy and neurophysiological structure of organisms, and the "synthetic method". The latter incorporates "general principles" that apply to both living organisms and the machines, considering both as complex adaptive systems. Craik, however, also indicated the risks implicit in the method, by emphasizing that models could be reduced to mere imitations of the phenomenon, i.e. experiments lacking any scientific interest as regards the explanation of the behaviour of organisms, because they did not share with the organisms any common functional principle. In fact, even Rosenblueth and Wiener (1945) point out that the artifacts could be relevant for the explanation in cognitive science only if they were instantiations of a "theoretical model", which ensured the basis for comparison between the natural and the artificial system (see also Tamburrini, Datteri 2005).The idea of the model as a test for theories is crucial here, together with the explicit formulation of the notion of a theory-model methodological cycle that would become pervasive in the next step of the evolution of the synthetic method.

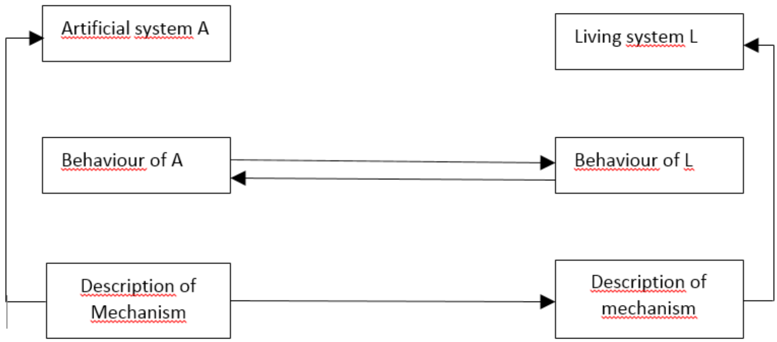


Figure 1. Methodological diagram for the analysis of simulative studies. Adapted from Datteri 2012.

This diagram (see Figure 1) describes the work of the researcher: she observes a biological behaviour and identifies a particular cognitive mechanism that may generate that behaviour (Figure 1, right side). She then builds an artificial simulation, or a robot, of the hypothesis of the mechanism under evaluation; she compares the behaviour of the simulation with the behaviour of natural agents; behavioural concordances and discrepancies are considered as the empirical basis for accepting or rejecting the hypothesis, under the assumption that the system has thoroughly simulated the hypothesis (see Datteri 2012, Grasso *et al*. 2000). Under a variety of epistemological and methodological assumptions (Webb 2008; Datteri, Tamburrini 2007), the synthetic method may therefore be helpful in identifying the mechanism underlying a particular (observed) behaviour. This may be called a *model-oriented* use of simulations (Datteri, Laudisa 2016).

Currently, the ambition to build machines able to reproduce the exact biological mechanisms faithfully are driven to greater and greater levels of biological accuracy. However, nowadays simulations are mostly used for another, rather different, purpose: they are essential to obtain fine-grained descriptions simply because there are no alternative ways to observe them at the same level of detail. The "molecular-level simulations" – as argued by Datteri and Laudisa (2016) – «are used as "computational microscopes" (Dror *et al*. 2012) to predict the behaviour of ion channels, under a variety of physiological conditions» (p. 26). Furthermore, in "evolutionary biorobotics" simulations are useful to obtain a desidered behaviour under a variety of conditions, reproducing the sensory-motor mechanisms and the physical structure of extinct animals (Long 2012; Plebe, Grasso 2016). The purpose of these studies is to obtain data on the behaviour of a system which is hard or impossible to observe through more conventional techniques, not to discover the mechanism underlying a particular behaviour. Simulations, in fact, are used to mimic conditions that are not empirically testable. This use of simulation can be called *data-oriented* (see also Boone, Piccinini 2016).

However, the main exponents of contemporary large-scale brain simulation projects are often ambiguous as to whether their goals are on the *data-oriented* or on the *model-oriented* side. As Roysam *et al.* (2009) stated: «Such simulations will offer more precise methods for testing potential biotechnology solutions to brain disorders, such as drugs or neural implants» (p. 2). Here they suggest that a «computer model of the brain could assist in discovering how brain behaviour would change in particular conditions, that is to say, under the effect of certain drugs or after connection with additional devices. This goal is closer to the data-oriented side, as is the goal of obtaining data on the target system» (Datteri, Laudisa 2016, p. 28). Similar ambiguities can be found in Eliasmith and Trujillo (2014) claiming that one of the reasons to build large-scale simulations is to understand unexplained brain disorders, such as autism and addiction. This achievement is closer to the *model-oriented* side, because it consists in the construction of theoretical models. Similarly, Kandel *et al*. (2013) point out that «the […] goal of these […] projects is to gain a better understanding of the anatomical, molecular and circuit bases for the logical operations carried out by the human brain» (p. 659). In their opinion, the Blue Brain Project «aims to understand the human brain by simulating its functions through the use of supercomputers» (*ibidem*). According again to Eliasmith and Trujillo (2014), another goal of large-scale simulations is «to develop and test new kinds of medical interventions, be they drugs or stimulation» (p. 3). As we have seen, obtaining data on the behaviour of a system which is hard or impossible to observe through more conventional techniques would be a *data-oriented* use of simulations.

Eliasmith and Trujillo (2014) also point out that a major purpose of building large-scale brain simulations is «to provide a way to organize and unify the massive amounts of data generated by the neurosciences» (*ibidem*). But, as we will see, whether and how simulations can really assist in integrating knowledge of the brain may depend on the use of a *model-oriented* method.

Cognitive science, and particularly modelling through simulations, makes use of two methodological families. These two methodologies seemingly have different goals: while *model-oriented* simulations go in search of the mechanism that might explain a cognitive process, *data-oriented* simulations seek to obtain data unavailable through alternative investigative strategies. Is it possible to integrate the two methods in order to study a cognitive process?

Levels of explanation in simulative methodology

The two kinds of simulation differ one from the other in the nature of their goals. This difference also reflects methodological differences. According to Datteri and Laudisa (2016), the first difference concerns the comparison between the behaviours of the artificial and the natural system. As previously seen, testing for *model-oriented* simulations is precisely such a comparison. Nevertheless, it «is not part of the *data-oriented* methodology, exactly because there is no data of the natural system to compare with the artificial behaviours» (p. 27).

A second difference concerns the degree of corroboration of the simulated mechanism. In *model-oriented* simulation studies, the artificial system must accurately simulate the model under scrutiny – otherwise, there would be no reason to bring the behaviour of an artificial system to bear on the plausibility of a natural model. Instead, to make a proper *data-oriented* use of a simulation, one has to assume that the model is a good model, otherwise there are no reasons to consider the behaviour of the simulation as the behaviour that the target system would have produced under such conditions.

In this conclusive section, we will propose two different kinds of integration between *data* and *model-oriented* simulations, in order to provide a way to use the *data oriented* simulation to understanding the cognitive mechanisms.

The first integration can be defined as "vertical". It refers to the possibility of inserting a *data-oriented* simulation in the methodological cycle (Figure 1). One of the goals of *data-oriented* simulations was to obtain data unavailable under normal circumstances. Consider again the methodological cycle presented in Figure 1. We may use the *data-oriented* simulations to produce data, thus forming the experimental basis of the *model-oriented* simulation. In other words, the *data-oriented* simulations could be exploited to generate, in a principled way, data to be fed to a *model-oriented* simulation, which eventually tries to explain the mechanism underlying the obtained data. In this case, the natural system can be replaced by the artificial system in the methodological cycle (Figure 2, right side).

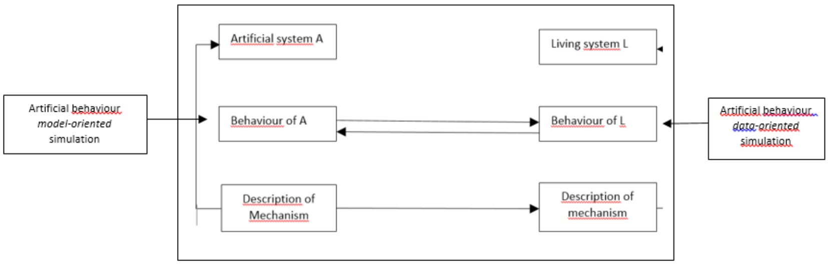


Figure 2. Methodological cycle (Datteri 2012) with two integrations. Vertical integration on the right and horizontal integration on the left.

The *data-oriented* simulation can also play an important role in favoring greater biological accuracy. The construction of supercomputers that simulate networks composed of billions of neurons is in itself an important technological advance. However, it is not obvious that the increase of the size of a simulation is interesting with respect to the integration or the prediction of the behaviour of the brain. According to the synthetic method, an artificial simulation would capture all and only the causally relevant factors for the production of the investigated behaviour (Boccignone, Cordeschi 2012). One of the unresolved problems of the synthetic method is underdetermination. It is possible, at least in principle, to generate material (simulation) models equivalent in performance (two machines will have the same performance). How, then, may we choose the one that really explains the investigated cognitive mechanism? "Horizontal" integration could respond to such question. As shown in Figure 2 (left side) the *data-oriented* simulations could serve as the simulative counterpart, in the case of underdetermination, in order to verify the working hypothesis underlying alternative simulations (*model-oriented*). In this case, if a *data-oriented* simulation is comparable, in performance, to the *model-oriented* simulation, the researcher should be encouraged to make the cognitive mechanism more explicit and to compare the underlying algorithmic processes with those of a *data-oriented* simulation. However, this does not imply that a simulation with greater biological accuracy is, necessarily, more suitable to explain the investigated cognitive process. In brief, the integration of the two methodologies seems possible even if the goal were *understand me* and not only *duplicate me*.

References

Arbib M.A. 2003, *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge.

Boccignone G., Cordeschi R. 2012, *Predictive Brains: Forethought and the Levels of Explanation,* in «Frontiers in Psychology», 6, 213, pp. 1-2.

Boone W., Piccinini G. 2016, *The Cognitive Neuroscience Revolution,* in «Synthese», 193(5), pp. 1509-1534.

Cordeschi R. 2000, *Early-Connectionism Machines*, in «AI & Society», 14(3), pp. 314-330.

Cordeschi R. 2002, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Kluwer, Dordrecht.

Cordeschi R. 2008, *Il Metodo Sintetico: Problemi Epistemologici nella Scienza Cognitiva*, in «Sistemi intelligenti», 20, (2), pp. 167-191.

Craik K.J.W. 1943, *The Nature of Explanation*, Cambridge University Press, Cambridge.

Datteri e. 2012, *Filosofia delle scienze cognitive: spiegazione, previsione, simulazione,* Carocci, Roma.

Datteri E., Laudisa F. 2016, *Large-scale Simulations of Brain Mechanisms: Beyond the Synthetic Method*, in «Paradigmi. Rivista di Critica Filosofica», 3, pp. 26-42.

Datteri E., Tamburrini G. 2007, *Biorobotic Experiments for the Discovery of Biological Mechanisms*, in «Philosophy of Science», 74(3), pp. 409-430.

De Garis H. *et al.* 2010. *A World Survey of Artificial Brain Projects, Part I: Large-scale Brain Simulations*, in «Neurocomputing», 74(1-3), pp. 3-29.

Dror R.O. *et al.* 2012, *Biomolecular Simulation: a Computational Microscope for Molecular Biology*, in «Annual Review of Biophysics», 41, pp. 429-452.

Eliasmith C., Trujillo O. 2014, *The Use and Abuse of Large-scale Brain Models*, in «Current Opinion in Neurobiology», 25, pp. 1-6.

Grasso F.G., 2000, *Biomimetic Robot Lobster Performs Chemo-orientation in Turbulence Using a Pair of Spatially Separated Sensors: Progress and Challenges,* in «Robotics and Autonomous Systems», 30, pp. 115-131.

Kandel E.R., *et al.* 2013, *Neuroscience Thinks Big (and Collaboratively)*, in «Nature Reviews Neuroscience», 14, pp. 659-664.
Long J. 2012, *Darwin's Devices. What Evolving Robots Can Teach us about the History of Life and the Future of Technology*, Basic Books, New York.
Markram H. 2006, *The Blue Brain Project*, in « Nature Reviews Neuroscience», 7, pp. 153-160.
Oudeyer P.Y. 2010, *On the Impact of Robotics in Behavioural and Cognitive Sciences: from Insect Navigation to Human Cognitive Development*, in «IEEE Transactions on Autonomous Mental Development», 2, pp. 2-16.
Plebe A., Grasso G. 2016, *The Brain in Silicon*: *History, and Skepticism*, in *Proceedings of the Third International Conference on History and Philosophy of Computing*, ed. by F. Gadducci, M. Tavosanis, pp. 273-286.
Rosenblueth A., Wiener N. 1945, *The Role of Models in Science*, in «Philosophy of Science», 12(4), pp. 316-321.
Rosenblueth A., Wiener N., Bigelow J. 1943, *Behaviour, Purpose and Teleology*, in «Philosophy of Science», 10, pp. 18-24.
Roysam B., Shain W., Ascoli G.A. 2009, *The Central Role of Neuroinformatics in the National Academy of Engineering's Grandest Challenge: Reverse Engineer the Brain*, in «Neuroinformatics», 7(1), pp. 1-5.
Russell S.B. 1913, *A Practical Device to Simulate the Working of Nervous Discharges*, in «Journal of Animal Behaviour», 3, pp. 15-35.
Santucci V.G., Cilia N.D., Pezzulo G. 2016, *The Status of the Simulative Method in Cognitive Science - the Current Debate and Future Prospects,* in «Paradigmi. Rivista di Critica Filosofica», 3, pp 51-74.
Simon H.A. 1996, *The Sciences of the Artificial*, MIT Press, Cambridge MA.
Tamburrini G., Datteri E. 2005, *Machine Experiments and Theoretical Modelling: from Cybernetic Methodology to Neuro-Robotics*, in «Minds and Machines», 15(3), pp. 335-358.
Webb B. 2008, *Using Robots to Understand Animal Behaviour*, in «Study of Behaviour», 38, pp. 1-58.

Take Another Little Piece of my Heart: A Note on Bridging Cognition and Emotions

Giuseppe Boccignone

Science urges philosophy to be more empirical and philosophy urges science to be more reflective. This markedly occurred along the "discovery of the artificial" (Cordeschi 2002): in the early days of Cybernetics and Artificial Intelligence (AI) researchers aimed at making machines more cognizant while setting up a framework to better understand human intelligence.

By and large, those genuine goals still hold today, whereas AI has become more concerned with specific aspects of intelligence, such as (machine) learning, reasoning, vision, and action. As a matter of fact, the field suffers from a chasm between two formerly integrated aspects. One is the engineering endeavour involving the development of tools, e.g., autonomous systems for driving cars as well as software for semantic information retrieval. The other is the philosophical debate that tries to answer questions concerning the nature of intelligence. Bridging these two levels can indeed be crucial in developing a deeper understanding of minds.

An opportunity might be offered by the cogent theme of emotions. Traditionally, computer science, psychological and philosophical research have been compelled to investigate mental processes that do not involve mood, emotions and feelings, in spite of Simon's early caveat (Simon 1967) that a general theory of cognition must incorporate the influence of emotion.

Given recent neurobiological findings and technological advances, the time is ripe to seriously weigh this promising, albeit controversial, opportunity.

In the heart of cognition

Affective neuroscience (Dalgleish *et al.* 2009, pp. 355-368) is helping us understand the neural circuitry that underlies emotional experience. It integrates functional neuroimaging, behavioural experiments, electrophysiological recordings, animal and human lesion studies and behavioural experiments striving to understand emotion at the neurobiological and psychological levels. Conflicting explanations at the psychological level, e.g. basic emotions vs. appraisal theories, find a novel synthesis at the neurobiological level. One outstanding example is Damasio's work (Damasio 1994, 1999).

Taking stock of such results, affective computing (Picard 2000) is dealing with artificial agents that aim at instantiating the ability to 1) recognize emotion, 2) express emotion, 3) "have emotions", the latter being the hardest task. So far, most current research focuses on 1) and 2), whereby machine learning-based affect detection plays a prominent role (Calvo, D'Mello 2010). In order to provide a thorough discussion of these aspects, we start by making clear the modelling strategy we adopt from now on.

Given a system (human observer) and its behaviour (e.g. facial expression, gaze shifts), together with available knowledge (psychological/neurobiological theories and descriptions, experiments and measurements), we set up a computational theory (Marr 1982) formalized in terms of Bayesian theory (Boccignone, Cordeschi 2007). More precisely, we draw on the Bayesian framework of Probabilistic Graphical Models (PGM) (Bishop 2006). In brief, by exploiting knowledge and constraints available both at the psychological and at the neurobiological description levels: 1) we identify the essential random variables (RVs) that ground the probabilistic model; 2) we encode the statistical dependencies between RVs in the PGM structure.

Coming back to the issue of dealing with the problem of affective expression generation/detection, one modelling example is presented in Figure 1. The time-varying RVs, $E(t)$ and $\mathrm{F}(t)$ stand for the latent affective state of an agent and the facial expression induced by such state at time t, respectively; such RV's are represented by graph nodes. The structural dependency (arrow) $E(t) \rightarrow \mathrm{F}(t)$ captures the statistical dependency of $\mathrm{F}(t)$ on $E(t)$, quantified via the conditional probability $P(\mathrm{F}(t) \mid E(t))$. The model is *generative*: $P(\mathrm{F}(t) \mid E(t))$ specifies the likelihood of generating an expression (by sampling) under a given affective state. The recognition problem (inferring the most plausible affective state given an observed facial expression), boils down to "inverting the arrows" by computing the posterior probability $P(E(t) \mid \mathrm{F}(t))$ via Bayes' rule. If $E(t)$ spans a discrete affective state-space, the model can account for a large number of computational models based on discrete theories á la Ekman (Ekman 1993, p. 384); if the space is continuous (e.g., specified via valence/arousal dimensions), it is suitable to cope with Russell's core affect theory (Russell 2003, p. 145).

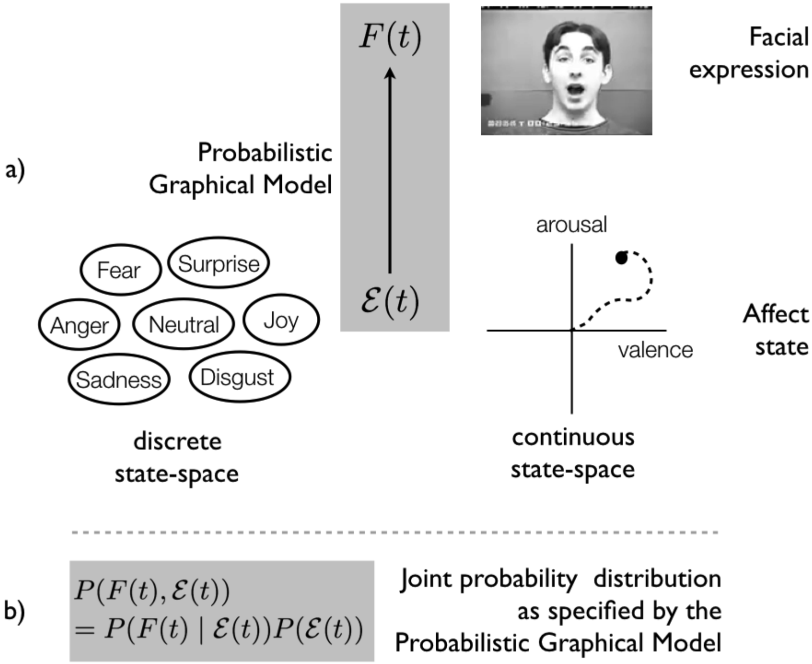


Figure 1. Modelling affective facial expression: a) the PGM models the dependency of the observed expression under the current affect state (discrete or continuous); b) the model represents a possible factorization of the joint probability P(F(t), E(t)).

Yet, the actual challenge is designing artificial agents that "have emotion" and use it for making decisions. Indeed, neurological studies indicate that decision-making without emotion can be impaired. Damasio's findings point to such an essential role of emotions (Damasio 1994).

It goes without saying, modelling emotion at the most general level is a mind-blowing endeavour for current research. Thus, we will focus on the integration of emotion with cognitive behaviour (Pessoa 2008) by drawing on the minimalist case of active sensing.

Where to look next?

Among the variety of active sensing behaviours, oculomotor behaviour (saccades, pursuit, fixational movements) is the least energy process. Though minimal, from a theoretical standpoint, a gaze shift action can be considered as the result of a decision-making process (conscious or unconscious) (Yang *et al.* 2016).

Such process can be modelled through the perception-action loop as a dynamic PGM (unfolded in time, top of Figure 2): *I(t)* denotes the stimulus, e.g. a time varying scene, and r*F (t)* is the point of gaze (center of the fovea) at time *t*; *A(t)* is the ensemble of RVs defining the oculomotor action setting (e.g., maintain current fixation or saccade in a certain direction); *W(t)* stands for the ensemble of RVs (e.g., features, objects) characterising the scene as actively perceived by gazing the stimulus at point r*F* ; *G* summarizes the given goal (e.g, search for a kid). The action setting dynamics *A(t) → A(t + 1)* and the scene perception dynamics *W(t) → W(t + 1)* are intertwined with one another through the gaze shift r*F (t)* → r*F (t + 1)*. The actual shift is recovered as the statistical decision of selecting a particular gaze location with probability *P(*r*F (t + 1) | A(t),W(t),* r*F (t)),* so to maximize the expected payoff under *G*, the current goal[1].

Perceptual decision-making calls for the notions of value and reward (Yang *et al.* 2016) that, in turn, pave the way to bringing emotions into the loop (Boccignone 2016). At the neurobiological level, it has been made clear that, crucially, cognitive (perceptual) and emotional contributions cannot be separated (Pessoa 2008), as outlined in Figure 2 (centre).

Indeed, there is a large body of evidence that responses from visual cortex reflecting stimulus significance are the result of simultaneous top-down modulation from fronto-parietal attentional regions and emotional modulation from the amygdala and the posterior orbitofrontal cortex (OFC). The affective value attributed to a stimulus – either consciously or unconsciously – drives attention and enhances the processing of emotionally modulated information (much like the physical salience of the stimulus), while exogenously driven attention influences the outcome of affectively significant stimuli (*ibidem*). At the same time, the cognitive control system (lateral prefrontal cortex, LPFC, anterior cingulate cortex, ACC) guides behaviour while handling goal-related information; action strategies incorporate val-

[1] For different instantiations of the model, see Clavelli *et al.* 2014, and Napoletano *et al.* 2015.

ue through the mediation of the nucleus accumbens, the amygdala, and the OFC. Basal forebrain cholinergic neurons provide regulation of arousal and attention while dopamine neurons located in the ventral tegmental area (vTA) modulate the prediction and expectation of future rewards (*ibidem*).

It is worth noting that neurobiological evidence is relevant for modelling purposes, if we surmise a certain degree of association between the neurobiological and the behavioural levels. We will further comment on this point in the final section of this note, but briefly, we assume that processes that support behaviour are implemented by the interaction of multiple areas (networks), which are dynamically recruited into multi-region assemblies (no "necessary and sufficient" brain regions). In this perspective, we draw on Damasio's cleavage between emotions and feelings, which are the first person experience of the corresponding emotion (Damasio 1994). An emotion is a neural reaction to a certain stimulus, realised by a complex ensemble of neural activations in the brain (internal emotional state). The latter often are preparations for (muscular, visceral) actions (facial expressions, heart rate increase, etc.), as a consequence the body will be modified into an "observable" emotional body state. Thus, we introduce the RV F*(t)* standing for visceral responses (e.g., heart rate, dermal response) that can be gauged via physiological measurement (ECG, skin conductance, etc.). Note in Figure 2 (centre panel) the central role of the amygdala and the OFC. Their tight interaction provides a suitable ground (Salzman, Fusi 2010, p. 173) for representing, at the psychological level, the core affect dimensions (Russell 2003). Core affect can then be functionally modelled as a latent space (Vitale *et al.* 2014) – see Figure 1 – spanned by $E(t)$. In addition, $C(t)$ indexes a higher cognitive level of interest. As a result, the original PGM is modified in the PGM shown at the bottom of the same figure.

Due to limitation of space, we are not entering details about software/hardware implementations. As to algorithms, a viable solution to provide a simulation of the model is that of exploiting the huge number of state-of-the-art machine learning algorithms (Bishop 2006). Learning and inference on PGMs can then be accomplished either through approximate optimization-based techniques (e.g., Variational Bayes) or stochastic techniques (Monte Carlo). Eventually, notice that, in the last decade, the number of public repositories has grown larger, where behavioural data gathered in realistic, natural setting experiments have been recorded by multiple modalities (Calvo, D'Mello 2010). Such data can be readily employed for model learning and validation.

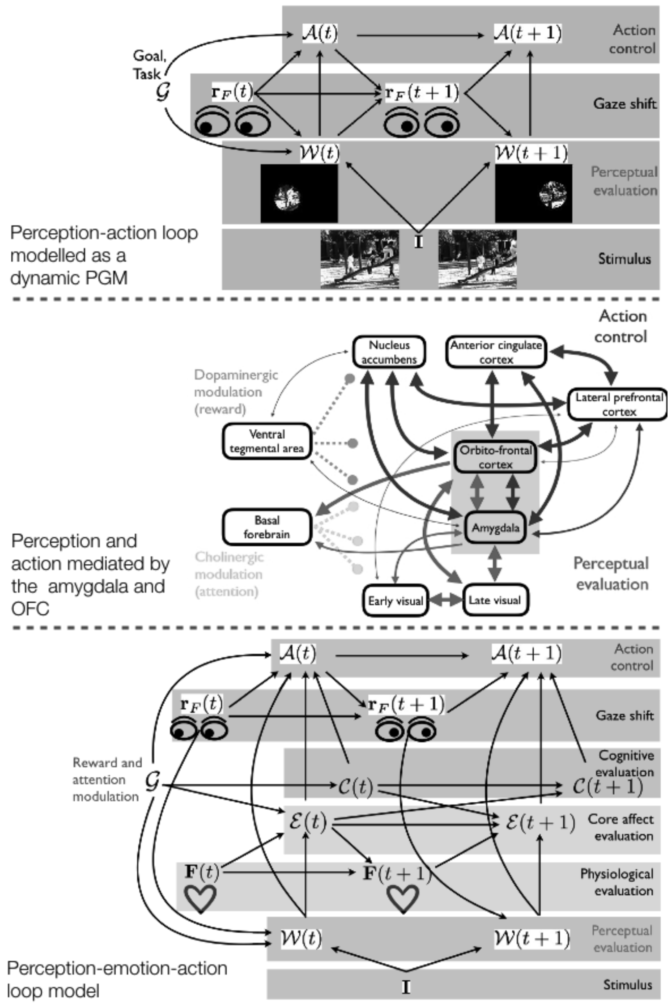


Figure 2. Top: the dynamic PGM of an active sensing loop. Centre: structural/functional constraints implied by circuits for visual processing and executive control. Bottom: the perception-emotion-action PGM.

Caveats on methodology

Though minimal, would the implemented version of the model meet the requirement of making decisions by virtue of "having emotions"? Such question entails a number of hindrances.

First, we have assumed, based on Damasio's distinction between emotions and feelings, to rule out the latter. Under such disentanglement, emotions are likely to be amenable to third person description (and thus modelled), whilst feelings would necessarily involve first person experience (opening to the conundrum of consciousness) (Trautteur 2016).

Second, we have set up a computational theory (Marr 1982) in the Bayesian framework of Probabilistic Graphical Models, where the RVs capturing essential behavioural properties are shaped in the PGM structure by using structural constraints suggested at the neurobiological level. Once implemented, the model is in principle suitable to simulate attentive behaviour conditioned by emotion. How things stand, putting the simulation of the model into work[2] is nothing but an instance of the *synthetic method* (Cordeschi 2002), i.e. the building of artefacts as explanatory models of living organisms. The synthetic method, *per se*, entails a variety of problems (Cordeschi 2008).

In particular, the "underdetermination" problem involves the choice of "the right grain of analysis for models". To handle the computational explanation at different grains/levels (Figure 3), we have adopted a revised form of Marr's framework (Marr 1982). In a Bayesian formalism (Boccignone, Cordeschi 2007), Marr's three-fold hierarchy can be re-organized into two levels (Knill *et al.* 1996): the *computational theory level*, which can be formalized precisely in terms of Bayesian theory, and the *implementation theory level* embedding both algorithmic and realization levels.

[2] For a nice discussion of simulation in cognition, see Santucci *et al.* 2016.

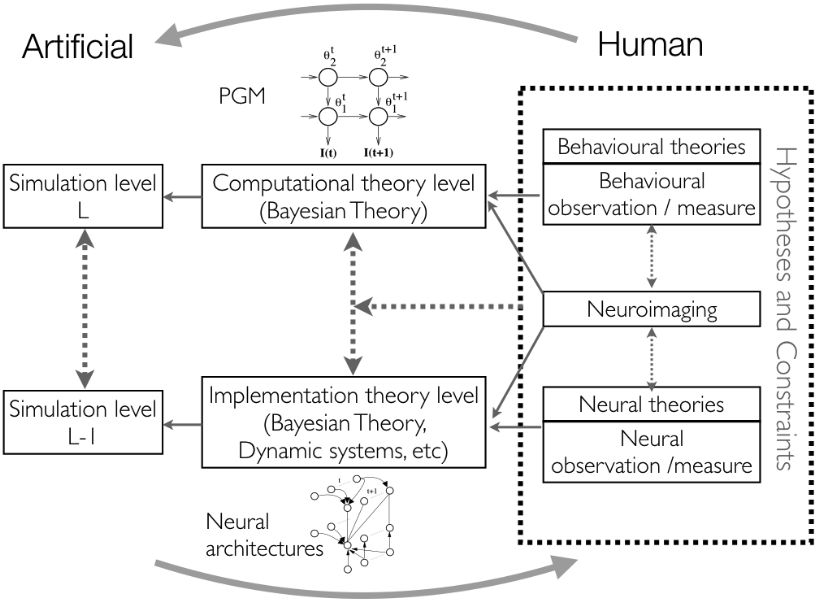


Figure 3. The synthetic method loop embedding different levels of explanation.

Differently from Marr (Boccignone, Cordeschi 2007), the notion of architecture becomes crucial: the PGM embodies constraints assumed by the scientist for his own purpose at the chosen level of explanation. The algorithmic level does not so far provide an autonomous level, rather one encompassing simulations of different grains (a coarse-grained simulation of Bayesian inference at the behavioural level, a fine-grained simulation at the neural level, see Figure 3).

As to the realization level, it is current practice to choose some formal neuronal model to the end (e.g., integrate-and-fire neurons, stochastic differential equations. simple binary, on/off models). This justifies the term "implementation theory": the realization level is but another kind of theoretical model (Abbott, Kepler 1990). Yet, further levels of reduction, and further theoretical models too, could be achieved (*ibidem*) going down in the hierarchy.

In summary, the Bayesian approach provides a sound formalization of Marr's functionalist intuition of a computational theory level. However, a deceptively simple question arises (Sprevak 2016, p. 92): how should we interpret Bayesian models? One option is *instrumentalism*, where Bayesian machinery should be understood as a formal device (*ibidem*) to describe human behavioural patterns concisely and to make predictions. The alternative is *realism*: models pick out real entities and processes in the human brain ("Bayesian brain hypothesis"; see Knill, Pouget 2004).

The Bayesian approach is advocated for handling uncertainty, stemming from lack of knowledge and from randomness. Going down in the explanation hierarchy, basic sources of randomness are classical dynamics unpredictability and quantum processes, which in living systems are likely to take place simultaneously and affect each other. Further, different levels of organization make things worse: multi-level interactions induce subsequent forms of randomness (Buiatti, Longo 2013). If one assumes a realistic stance, these "living matter" effects pose serious challenges to the functionalism captured by the computationalist account (Cordeschi, Frixione 2007). Indeed, there is severe criticism (Longo 2009) in the ability of digital computation to fully reproduce (not just mimicking) this dynamics even in simple cases (deterministic unpredictability).

A viable shortcut (Cordeschi, Frixione 2007) is the "encapsulation" of any critical level dealing with a non-Turing computable function in an embedded subsystem, so to consider only the computable outputs that might be relevant for higher embedding levels. However, even the "encapsulation" practice is not, at least in principle, unquestionable. Since minor changes in one level might be amplified by the exchanges with another level, such approach might rule out underpinning properties at the biological level, crucial for the overall behaviour of the system, especially in the case of emotions. Even discarding the conundrum of feelings, yet emotions use both neural and humoral routes, so that the resulting emotional state involves continuous, analogue changes within the body proper, e.g., viscera, internal milieu, etc.

All the above issues let us surmise that there is much work left for scientists. And even more for philosophers.

References

Abbott L.F., Kepler T.B 1990, *Model Neurons: From Hodgkin-Huxley to Hopfield*, in *Statistical mechanics of neural networks - Lecture Notes in Physics*, 368, ed. by L. Garrido, Springer-Verlag, Berlin-Heidelberg, pp. 5-18.

Bishop C.M. 2006, *Pattern Recognition and Machine Learning*, Springer, New York.

Boccignone G. 2016, *A Probabilistic Tour of Visual Attention and Gaze Shift Computational Models*, in *International Workshop "Vision over Vision: Man, Monkey, Machines, and Network Models"*, Osaka (Available at http://arxiv.org/abs/1607.01232).

Boccignone G., Cordeschi R. 2007, *Bayesian Models and Simulations in Cognitive Science*, in *Models and Simulations 2*, Tilburg, PhilSci-Archive (Available at http://philsci-archive.pitt.edu/3556/).

Buiatti M., Longo G. 2013, *Randomness and Multilevel Interactions in Biology*, in «Theory in Biosciences», 132(3), pp. 139-158.

Calvo R.A., D'Mello S. 2010, *Affect Detection: An Interdisciplinary Review of Models, Methods, and their Applications*, in «IEEE Transactions on Affective Computing», 1(1), pp. 18-37.

Clavelli A., Karatzas D., Llados J., Ferraro M., Boccignone G. 2014, *Modelling Task-Dependent Eye Guidance to Objects in Pictures*, in «Cognitive Computation», 6(3), pp. 558-584.

Cordeschi R. 2002, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Kluwer Academic Publishers, Dordrecht.

Cordeschi R. 2008, *Steps toward the Synthetic Method*, in *The Mechanical Mind in History*, MIT Press, Cambridge, pp. 219-258.

Cordeschi R., Frixione M. 2007, *Computationalism Under Attack*, in *Cartographies of the Mind: Philosophy and Psychology in Intersection*, ed. by M. Marraffa, M. De Caro, F. Ferretti, Springer Netherlands, Dordrecht, pp. 37-49.

Dalgleish T., Dunn B.D., Mobbs D. 2009, *Affective Neuroscience: Past, Present, and Future*, in «Emotion Review», 1(4), pp. 355-368.

Damasio A. 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam, New York.

Damasio A. 1999, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace & Company, New York.

Ekman P. 1993, *Facial Expression and Emotion*, in «American Psychologist», 48(4), pp. 384-392.

Knill D.C., Pouget A. 2004, *The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation*, in «Trends in Neurosciences», 27(12), pp. 712-719.

Knill D.C., Kersten D., Yuille A. 1996, *Introduction: A Bayesian Formulation of Visual Perception*, in *Perception as Bayesian Inference*, ed. by D.C. Knill, W. Richards, Cambridge University Press, Cambridge (UK), pp. 1-21.
Longo G. 2009, *Critique of Computational Reason in the Natural Sciences*, in *Fundamental Concepts in Computer Science*, ed. by E. Gelenbe, J.-P. Kahane, Imperial College Press, London, pp. 43-70.
Marr D. 1982, *Vision*, W.H. Freeman, New York.
Napoletano P., Boccignone G., Tisato F. 2015, *Attentive Monitoring of Multiple Video Streams driven by a Bayesian Foraging Strategy*, in «IEEE Transactions on Image Processing», 24(11), pp. 3266-3281.
Pessoa L. 2008, *On the Relationship between Emotion and Cognition,* in «Nature Reviews Neuroscience» 9(2), pp. 148-158.
Picard R.W. 2000, *Affective Computing*, MIT Press, Cambridge.
Russell J.A. 2003, *Core Affect and the Psychological Construction of Emotion*, in «Psychological review», 110(1), pp. 145-172.
Salzman C.D., Fusi S. 2010, *Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex*, in «Annual Review of Neuroscience», 33, pp. 173-202.
Santucci V.G., Cilia N.D., Pezzulo G. 2016, *The Status of the Simulative Method in Cognitive Science: Current Debates and Future Prospects*, in «Paradigmi», 3, pp. 47-66.
Simon H.A. 1967, *Motivational and Emotional Controls of Cognition*, in «Psychological review», 74(1), pp. 1193-1216.
Sprevak M. 2016, *Philosophy of the Psychological and Cognitive Sciences*, in The Oxford Handbook of Philosophy of Science, ed. by P. Humphreys, Oxford University Press, Oxford.
Trautteur G. 2016, *Consciousness, Illusory Freedom, Double Feel,* in «Paradigmi», 3, pp. 12-22.
Vitale J., Williams M-A., Johnston B., Boccignone G. 2014, *Affective Facial Expression Processing via Simulation: A Probabilistic Model*, in «Biologically Inspired Cognitive Architectures», 10, pp. 30-41.
Yang S.CH., Wolpert D.M., Lengyel M. 2016, *Theoretical Perspectives on Active Sensing*, in «Current Opinion in Behavioural Sciences», 11, pp. 100-108.

The Role of Emotions and Intrinsic Motivations on Decision-Making: a Comparison between Natural and Artificial Systems

Valentina Trombetta

Introduction

Decision-making is an important and multifaceted cognitive process. Every agent has to be able to evaluate the implications of his/her actions to ensure his/her welfare in a given environment. As such, decision-making pertains to both human beings and robots, because they live in a specific environment that demands choices to achieve specific goals, such as survival.

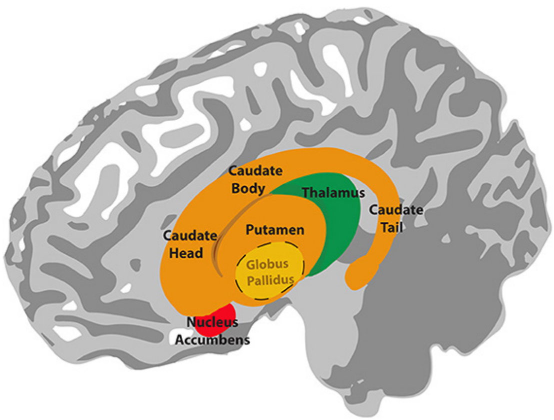


Figure 1. Basal ganglia structures. Source: Lim *et al*. 2014.

From a neurological point of view, the main areas of the brain involved in this process are the basal ganglia and the amygdalae. Neurological research shows that the basal ganglia and cortex are involved in decision-making between alternative actions (Bogacz, Gurney 2007). The basal ganglia are a large collection of subcortical structures and are part of the extrapyramidal motor system, the part of the motor system concerned with automatic movement. In addition to these functions, new data support an expanded role for the basal ganglia in cognitive processes (Berns, Sejnowski 1996).

On the other hand, the involvement of the amygdala is important, because it plays a primary role in emotional processes. Emotions can be considered a collection of neural and chemical responses, which have a twofold function: to produce a specific reaction to the inducing situation and to regulate internal states of the organism that produce specific reactions. Emotions can also influence the process of choice as judgement is engaged. Another psychological aspect that affects choice is intrinsic motivation, as it generally guides behaviour.

However, beyond neurobiology, decision-making is a cognitive process that involves issues coming from different fields of knowledge, like neuroscience, psychology and philosophy. In this paper, we focus on the psychological point of view, because the overlap[1] with robotics enables us to better understand this process through a comparison between natural and artificial systems.

Emotions

Emotions have a role in decision-making: they help reasoning, especially in personal and social matters. They guide us to choose the best course of action and to avoid the worst, in view of our general welfare. A simple example is the emotion of fear: if we are on a dark and isolated street, we can avoid possible dangers by changing course. Everything depends on our level of fear and concern.

The neuroscientist Antonio Damasio describes emotions as follows:

> Emotions are a complicated collection of chemical and neural responses, forming a pattern; […] emotions are about the life of an organism, its body to be precise, and their role is to assist the organism in maintaining life (Damasio 1999, p. 51).

[1] Several theories and experiments in robotics try to simulate in artificial systems the processes of emotions and intrinsic motivations.

Furthermore, he states that emotions are biologically determined, that they are produced in the subcortical regions and that they can be engaged automatically without consciousness, using the body as their theatre. Emotions take place when the body perceives new objects or events through its sensory devices, as well as when certain objects or events are recalled from memory. Different emotions are induced by different brain systems and the specific brain sites of interest are the following: brain-stem region, hypothalamus, basal forebrain, amygdala, sectors of the anterior cingulate region and parts of the ventromedial prefrontal region. For Damasio, there is a mutual and continuous relationship between the body and brain's states inducing emotions, which he calls "as if body loop". The first step of this loop is the engagement of the organism by an inducer of emotion. For example, a specific object visually processed or, in other words, a visual representation of that object. In the second step, the signals of the elaboration of the object's image activate all neural sites that are prepared to respond to the particular class of inducer to which the object belongs. Finally, in the third step, emotion induction sites send a number of signals towards other brain sites and towards the body. The combination of these three steps causes a momentary and appropriate emotion.

Damasio's work shows that emotion is integral to the processes of reasoning and decision-making, for better or for worse. An example of impairment caused by the absence of an emotion is one of his patients, suffering from Urbach-Wiethe disease[2], which brings about a total calcification of both amygdalae. While the patient maintains normal cognitive and social skills, her inability to recognize suspicious faces or situations makes her incapable of fear, and thus of avoiding social risks. This shows how an emotion can affect an action's selection and how emotions are important for survival.

There are many theories that devise strategies for computing the process of emotions[3]. One is in the *Component Process Model* by Klaus R. Scherer. In Scherer 2009, he claims:

> Emotion will be considered here as a bounded episode in the life of a system that is characterized as an emergent pattern of component synchronization,

[2] A rare autosomal recessive condition characterized by abnormal depositions of calcium in the skin and throat (Damasio 1999).

[3] For example, theoretical and practical developments about Affective Computing and Social Robots.

> preparing adaptive action tendencies to relevant events, as defined by their behavioural meaning and aiming at establishing control precedence over behaviour (Scherer 2009, p. 3459).

When there is a relevant event, the system evaluates it and its outcomes through a set of criteria, called SEC (*stimulus evaluation checks*)[4]. The result of this assessment is a motivational effect that changes the motivational state before the occurrence of the event. This new motivational effect brings changes in the automatic nervous system, in particular somatovisceral effects, and in the somatic nervous system, through motor expressions in face, voice and body. Furthermore, to compute procedures of synchronization non-linear dynamic systems techniques are required, which might be drawn from self-organization theory or chaos theory.

Both Damasio and Scherer point to the connection between the internal systems of an organism, his body and the environment to produce emotions. Relevant events in the world are appraised from specific brain sites, and then changes occur in different subsystems of the organism to represent, cognitively and physically, a particular emotion. On the basis of the emotion that is produced there will be a different action's selection between the available options:

> It is also important to note that while the biological machinery for emotions is largely present, the inducers are not part of the machinery, they are external to it. […] Emotion and the biological machinery underlying it are the obligate accompaniment of behaviour, conscious or not (Damasio 1999, pp. 57-58).

Intrinsic motivations

The other psychological aspect that can affect decision-making is motivation. Motivations, whether extrinsic or intrinsic, have three important functions: to guide behaviours in order to satisfy the most important needs, to establish the amount of energy required to perform action selection and to generate learning signals[5].

Extrinsic motivations are mechanisms, which guide the learning of skills and knowledge on the basis of homeostatic needs detected within the

---

[4] They are four: relevance, implications, coping potential and normative significance.

[5] Learning signals guide learning mechanisms to acquire behaviours that increase fitness.

visceral body. The learning signals produced by extrinsic motivations tend to disappear when the homeostatic needs are satisfied but resurface when the needs return. They are common in different species. Examples could include physiological needs, like hunger, thirst, sleep and so on.

Intrinsic motivations are mechanisms, which guide the learning of skills and knowledge on the basis of the levels and the variations of these skills and knowledge detected within the brain. The learning signals produced by intrinsic motivations tend to decrease when the skills and the knowledge that they induced are acquired. They emerged later than extrinsic motivations during evolution. An example is the curiosity that drives the desire to gain new knowledge, which is very important for cumulative learning, that is, the ability to acquire skills and knowledge that improve over time, to be exploited in new situations and environments.

For our purpose, intrinsic motivations are more relevant. We can divide them into *knowledge-based* and *competence-based* (Baldassarre 2011). The first one is based on what the system (natural or artificial) knows, meaning the capacity of the system to change its environment, its body as well as its interaction between body and environment. The second one is based on what the system can do. In other words, the capability of the system to have specific effects on the environment. It is important to have both of them to perform adaptive cumulative learning.

Generally, the extrinsic and intrinsic motivations allow an intelligent system to survive in its environment and this principle is valid also for artificial systems, through the use of reward in the learning's algorithm. Two cases will be illustrated below: the robot arm Katana by Hung Ngo (2012) and Curious Robot by Dario Di Nocera et al. (2014).

Ngo's experiment has been performed on both the simulated and the real robot arm Katana and uses only an intrinsic reward to guide the behaviour of the system. The setting is a blocks-world formed by eight different blocks with different colours and different heights. There is a camera to see and recognize the different blocks and when there are concealed blocks, their features and dispositions are stored in a memory module. It is implemented with an online learning algorithm and the intrinsic reward is artificial curiosity. The task of this experiment is divided into two steps: the first one is to choose a placement location, and the second is to choose which is the right block to put in that specific location. Furthermore, after the execution of the second step there is a query condition, which must indicate when an outcome is statistically known or unknown. This is important for a sequential decision process because it is based on the expected learning progress. The behaviour of the system works in an exploration-exploitation trade-off: it exploits the acquired knowledge to make the best choice and explores the predicted actions to improve its representation of the setting's model. The result is that the Progress-driven artificial curiosity is a more general method for balancing exploration and exploitation (Ngo 2012).

Figure 2. The Katana robot arm. Source: Ngo 2012.

The second experiment is Curious Robot (CR), so called by his author Di Nocera. In this case, extrinsic motivations as well as intrinsic motivations are simulated. There are three differently coloured blocks with corresponding functions: green blocks for the function EAT, blue blocks for the function DRINK, black blocks for the function AVOID. A threshold indicates a wellness state of the system. When the latter is under the threshold, the system needs to eat and drink. When the wellness state is over the threshold, it can explore its environment. The task of this experiment is to survive in this setting, with obstacles, for a given time. It is implemented with Q-learning[6] and the intrinsic reward is curiosity. There are two components to implement curiosity. The first one is the Wanting, the residual value of Energy for the robot body, this means that curiosity can only oc-

[6] Q-learning is a reinforcement algorithm to find an optimal action-selection policy.

cur if the wellness state is not critical and the robot can spend energy on behaviours not associated with primary needs. The second component is the Liking, the pleasure due to novel situations that is the level of Novelty in the exploration of the learning states. Finally, Di Nocera *et al.* (2014) compare CR with the behaviour of another robot: Not-Curious Robot (N-CR), so called because it does not have artificial curiosity. The result is that CR has safer behaviour than N-CR, as it eats and drinks more frequently and not only when the wellness state is under the threshold. Moreover, it has more time to explore, knows its environment and is more able to avoid the objects and walls to survive.

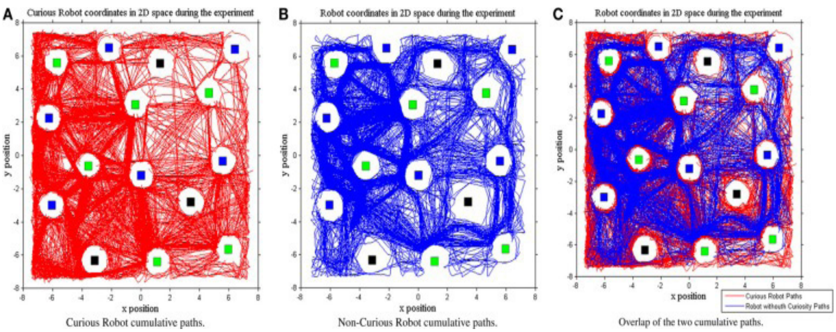


Figure 3. Experimental results for Curious (A-red line) or Non-Curious (B-blue line) Robots. These diagrams show the space of the environment explored. Source: Di Nocera *et al*. 2014.

These experiments show that motivations are important in an intelligent being's life, whether natural or artificial.

Conclusions

From a philosophical point of view, all these aspects can be discussed in a specific theoretical framework. In particular, in an enactive and dynamic approach. Thre term "Enactivism" was introduced in 1991 by Francisco J. Varela, Evan Thompson, and Eleanor Rosch in their book *The Embodied Mind*. In this approach, the cognitive agents are conceived as autonomous systems:

that is, self-organizing systems that generate and maintain themselves and enact[7] their cognitive domains. Furthermore, cognition is described like the exercise of skillful know-how in situated and embodied action.

The central idea of "Dynamicism" is that cognitive agents are dynamic systems and action, perception and cognition should be explained in dynamic terms, that is, as continuous coevolution[8]. This approach includes dynamic systems modelling (mathematical modelling of empirical systems), and experimental investigations of biological and psychological phenomena informed by these tools (Thompson 2007).

The fundamental point is that Enactivism and Dynamicism share an important idea: the relationship between brain, body and environment as the basic structure for the emergence of cognition. As we have seen, decision-making is a cognitive process affected by emotions and intrinsic motivations and both of them are possible only for agents in an environment. Indeed, emotions involve a biological aspect, brain and body, as well as an external aspect, because the stimuli that cause emotions are in the perceptual world of the agent. It is the same for intrinsic motivations, which include brain and body, but also an exchange of input and output with the outside that allows the agent to act.

The final question is: is it possible for artificial agents to simulate a complete decision-making? A complete decision-making requires an agent who chooses an action not optimal but the best for his survival, and not using only reason but also emotions and motivations. As we have seen earlier, artificial systems are able to simulate emotions and intrinsic motivations as elements steering their future behaviour. The basic principle about the deep connection between brain, body and environment seems to be respected in human beings as well as in robots. The fundamental difference is the way in which behaviour is expressed: robots do not have a biological body and brain, but specific algorithms can reproduce some particular processes. Robots have a different structure, which tends to simulate the living processes of natural systems. Until now, in the state-of-art robotics there are several models, each of which simulates a particular cognitive process (memory, emotion, and so

[7] Enaction means the action of enacting a law, but it also connotes more generally the performance or carrying out of an action (Thompson 2007).

[8] Cognition is seen as the flow of complex temporal structures mutually and simultaneously influencing each other (Thompson 2007).

on). This means that there is a lack of complexity such as it is required for a complete decision-making. Not only emotions or motivations, but both of them and other elements are necessary together. So, we should wait for the future developments in robotics to see if it is really possible to develop a complete decision-making in artificial systems.

References

Baldassarre G. 2011, *What are Intrinsic Motivations? A Biological Perspective*, in «Proceedings of the International Conference on Development and Learning and Epigenetic Robotics», IEEE, Piscataway NJ, pp. E1-8.

Berns G.S., Sejnowski T.J. 1996, *How the Basal Ganglia Make Decisions?*, in *Neurobiology of Decision-making*, ed. by A.R. Damasio, H. Damasio, Y. Christen, Springer, Berlin-Heidelberg, pp. 101-113.

Bogacz R., Gurney K. 2007, *The Basal Ganglia and Cortex Implement Optimal Decision Making between Alternative Actions,* in «Neural Computation», 19(2), pp. 442-477.

Damasio A. 1999, *The Feelings of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt, New York.

Di Nocera D., Finzi A., Rossi S., Staffa M. 2014, *The Role of Intrinsic Motivations in Attention Allocation and Shifting*, in «Frontiers in Psychology», 5:273.

Ngo H., Luciw M., Forster A., Schmidhuber J. 2012, *Learning Skills from Play: Artificial Curiosity on a Katana Robot Arm*, in «Proceedings of the International Joint Conference on Neural Networks», pp. 561-572.

Lim S-J., Fiez J.A., Holt L.L. 2014, *How May the Basal Ganglia Contribute to Auditory Categorization and Speech Perception?,* in «Frontiers in Neuroscience», 8:230.

Scherer K.R. 2009, *Emotion are Emergent Processes: They Require a Dynamic Computational Architecture*, in «Philosophical Transactions of the Royal Society B», 364(1535), pp. 3459-3474.

Thompson E. 2007, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, The Belknap Press of Harvard University Press, Cambridge-.

Varela F.J., Thompson E., Rosch E. 1991, *The Embodied Mind: Cognitive Science and Human Experience*, The MIT Press, Cambridge MA-London.

The Synthetic Approach and the Evolution of Cognition

Francesco Bianchini

Introduction: biology and cognition

The birth of artificial intelligence (AI) and the concurrent start of the logic/symbolic tradition in intelligent artificial systems in the 1950s and the 1960s led to the neglect of evolutionary issues as relevant for the study of cognition. This trend has been reversed in the last thirty years with the new AI and new cognitive science, that is, a cognitive science characterized by the 4E – embodied, embedded, enactive, extended – approach, within the general framework of embodied cognition and sensory-motor and environmental explanation of cognitive capabilities.

Such a new trend of cognitive science and AI, at least the AI devoted to the explanation of intelligent behaviour, is certainly more connected to biology than the earlier one. Neural networks and connectionism – brain inspired approaches –, artificial life, biologically inspired cognitive architectures, swarm cognition and situated robotics, embodied cognition and the sensory-motor explanation of cognitive capabilities are examples. However, even though one may think that biological heuristics in AI are akin to an evolutionary explanation of cognition, the equation is not so simple, a point which is usually not emphasized. Not all of biology is evolutionary, and the biological inspiration in AI and cognitive science needs therefore to be more precisely defined.

AI, biology and evolution

The relationship between biology and AI is old, older than AI itself. Cybernetics, one of the discipline that concurred to AI birth (Cordeschi 2002), deals with the problem of "Control and Communication in the Animal and the Machine" (Wiener 1948). In the 1940s and 1950s, the mechanistic explanation of the teleological behaviour typical of biological systems fostered a new approach to the study of the mind within a computational framework and of computer simulations of purpose-oriented autonomous behaviour. Moreover, the work of Turing on embryology and morphogenesis – the "biological" Turing (1948) – as well as the work of von Neumann on self-reproductive systems and self-replicating automata (von Neumann 1951) exploited biological ideas to build computer simulations of artificial autonomous systems, thus laying the foundations of artificial life. Both mathematicians worked before the officially recognized birth of AI in 1956, providing the basis for its conceptual framework.

While all these works are usually considered seminal for artificial life, it is only von Neumann's research that is strictly related to evolution. Actually, the main goal of these early research programs on self-regulating machines and self-replicating systems shared, notwithstanding the differences between Wiener, Turing and von Neumann, was the aim to explain and reproduce the self-organization at the basis of living systems.

On the other hand, self-organization is also the focus of artificial life since the 1980s, that is the study of the properties of life in general, at a more abstract level than other more "biological" disciplines, or, in the words of the founder of this field, life *as it could be* rather than life *as we know it* (Langton 1989). Computer simulation and modelling methodologies of artificial life are focused on reproducing the spontaneous emergence of order, starting from simple properties and the intrinsic structure of the system, independently of its interaction with the environment and without the intervention of a designer or a programmer (Boden 2006).

Artificial life, therefore, has to do with emergence and self-organization, but, differently from other approaches such as biomimetics or synthetic biology, is more interested in an abstract notion of life phenomena and biological aspects than in the building or reproduction of particular parts of a living creature. In this case yet, the use of evolutionary principles is just partial and limited.

The only work addressing entirely the principles and the mechanisms of evolution in the early AI and cognitive science period is that of von Neumann on self-replicating machines[1] or, in his own term, automata. Von Neumann's main aim was to catch the properties of self-reproducing and self-replicating systems (which are, actually, not the same thing). In the outline sketched in the Hixon Symposium of 1948 von Neumann speaks of a physical system capable of self-reproducing itself and equates the possible copy errors, with

[1] On this topic, see Bianchini 2016.

their heritability, to mutations in genetic evolution. It was only with the logical approach to these systems, thanks to the contribution of Stanislaw Ulam[2], however, that von Neumann was capable to give the foundations of tessellation structures or, as they have been called later, cellular automata.

Cellular automata are the main pathway by which evolution comes into the more recent field of AI and cognitive science. They are at the basis of the Game of Life by John Conway, which became soon part of the new artificial life field, bringing into it the logical mechanisms of evolutionary theory, as a cellular automaton is capable to self-replicate[3]. Another research program stemming from von Neumann's work on tessellations is the one by John Holland, which extended the cellular automata theory towards an even more evolutionary course. In fact, in Holland's works, "mutation" is not just a possibility. On the contrary, it is an integral part of the methodology he developed. His method is based on the idea that programs could be symbolized in the binary code, as in a Universal Turing Machine or a computer at the level of machine language, and they may be split and combined, with a certain amount of noise or copy errors, to generate new programs. Over the generations, the best programs, i.e. programs with the best performance in accordance with the expected results, are selected to generate new programs more and more adapted to the task to be accomplished. Adaptation and selection are notions Holland has taken especially from evolutionary theory. Holland's main contribution to AI and cognitive science in the mid-1970s (Holland 1975) is a general theory of "genetic algorithms" and a logical theory of adaptive systems and represents the official entry of evolution in such disciplines.

Genetic algorithms, even though they are not the first application of evolutionary principles to programming techniques, made the evolutionary approach very popular, and also feasible thanks to the increase of computer and processor power. Besides, they did contribute to the development of evolutionary computation, an AI subfield inspired by biological evolution and characterized by algorithms for global optimizations strategies, especially for problem solving[4]. Evolutionary computation has developed lately into a new subfield, evolutionary robotics[5], which simulates different generations of ro-

---

2 See Bialynicki-Birula, Bialynicka-Birula 2004.

3 In fact, not every cellular automaton has the lifelike property of self-replication.

4 See, for example, De Jong 2006.

5 Nolfi, Floreano 2000.

bots by evolving their control software and transferring it to the next generation of physical robots. This is also a way to reproduce the open-endedness of evolution in the physical world.

In her detailed history of AI and cognitive science Boden (2006) lists a series of fields that link together evolution, on one hand, and AI and cognition, on the other. In addition to the theories of the forerunners of AI (Wiener, Turing, von Neumann), she deals, of course, with evolutionary programming, genetic algorithms and evolutionary robotics. She mentions, however, also evolutionary psychology and evolutionary semantics, which both explain cognitive capabilities and mental contents by natural selection and evolutionary history. But in the case of these latter disciplines the major claim is rather that something is the result of evolution, which is quite different from exploiting the principles of evolutionary theory to explain and simulate cognitive capabilities and intelligent behaviour by "building" artificial autonomous systems. What is at stake here is rather to employ the power of evolution to carry out processes of transition from chaos to order, and the emergence of cognition, in a bottom up manner.

Evolution, the synthetic approach and a conclusion

Transition from chaos to order, that is from unorganized to organized systems, is typical of different AI approaches[6]. Neural networks and the connectionist approach (Rumelhart *et al.* 1986), especially in distributed and subsymbolic systems, are one of the main examples, but there are also other approaches: the subcognitive models (Hofstadter *et al.* 1995)[7], the swarm cognition approach (Trianni *et al.* 2011), among others. Even though all these approaches can be considered biologically inspired and aim at modelling adaptive systems, not all of them involve evolution and evolutionary methods.

It has been recently suggested that the synthetic method, or the method of model building, is a method shared both by old and new AI and cognitive science (Cordeschi 2008). The synthetic method has been historically

[6] For a very early discussion on this topic with reference to intelligent systems, see Turing 1948.

[7] In these models, evolutionary methods are just a part of the heuristics used to exploit the transition from randomness to a deterministic view (the solution) in a process of problem solving. Other bio-inspired heuristics are drawn from cellular metabolism and immune system; see Mitchell 2006.

defined as the search for the fundamental principles that allow an organism to learn and the engineering of mechanical artifacts according to these theoretical principles (Craik 1966). To this aim, the major issue is to build into the model, that is synthetize, the right constraints, i.e. those relevant for explaining a cognitive theory. In a unified view of AI and cognition studies, therefore, «what distinguishes old and new cognitive science is the choice of the level of explanation at which the right constraints for the models are to be introduced» (Cordeschi 2008, p. 245). Within the framework of the synthetic approach, cognition is a matter of constraints, because the simulative aspect requires model building, and models require the right i.e. relevant constraints to be considered the appropriate models for testing the explicative power of a cognitive theory. Without the right constraints, we have just imitation or mimicking without explicative power.

This is a very crucial epistemological problem for every science of the artificial committed to the explanation of a special phenomenon, even though many disciplines do not consider it so crucial. For example, synthetic biology is not committed to the explanation of life phenomena, but with the artificial construction or re-design of biological parts and systems for useful purposes (Freemont, Kitney 2012). Thus, it neglects the possible explanatory potential of biological synthesis to understanding life. Specific theories about living beings could provide new methods and purposes in designing and building synthetic biological parts. I do not want to claim that synthetic biology is an a-theoretical discipline. I just want to suggest that it misses the epistemological questions which arise from «the conspicuous difference between the production of artificial life and the artificial production of life» (Keller 2002, p. 19). With regard to AI and cognitive science, this may be tantamount to claim that there is a difference between the production of artificial cognition and the artificial production of cognition. This ambiguity is mirrored in the traditional split between the psychological AI and the engineering AI. While the former is more concerned with the explanation of cognitive phenomena, the latter is more concerned with the results achieved by employing useful techniques.

I think the same ambiguity appears with regard to evolution. Evolutionary methods exploit the power of selection and focus on adaptiveness with reference to a specific context (not necessarily a physical environment). Thus, the constraints in models using evolutionary methods are related to the interaction with that context. From a synthetic point of view, the transition from chaos to order, or from randomness to stability in response to a specific environment, which characterizes the evolutionary approaches in AI and cognition, is not easily traced back to a set of theoretical constraints that explain directly a cognitive capability. On one hand evolutionary method and techniques in AI appear to be more intended to achieving a system able to perform a class of specific tasks than committed to a clear explanatory goal; on the other, evolutionary methods allow to build systems capable to interact increasingly better with the real world – especially in evolutionary robotics – and that are not just models of something cognitive or autonomous or intelligent, but that are on their own, at the end of evolutionary process, something cognitive or autonomous or intelligent.

Evolutionary techniques are a large part of the more recent research in AI and cognition as well as in other synthetic discipline such as synthetic biology. New trends show a concurrence of interests in these discipline[8]. Evolution and evolutionary methods are, however, just a part of the new biologically-inspired approaches to AI and cognition, but with their peculiarities they allow to draw closer different disciplines with a confluence of interest in the biological world and, at the same time, to catch the power of life to attain cognition in the very way life produced cognition. Nevertheless, their explanatory power about cognition is the same one we could expect from the biological evolution in explaining cognition.

References

Bialynicki-Birula I., Bialynicka-Birula I. 2004, *Modeling Reality: How Computers Mirror Life*, Oxford University Press, Oxford.

Bianchini F. 2016, *Artificial Intelligence and Synthetic Biology: A Tri-temporal Contribution*, in «BioSystems», 148, pp. 32-39.

Boden M. 2006, *Mind as Machine. A History of Cognitive Science,* Oxford University Press, Oxford.

Copeland J. 2004, *The Essential Turing*, Clarendon Press, Oxford.

Cordeschi R. 2002, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Kluwer, Dordrecht.

Cordeschi R. 2008, *Steps Toward the Synthetic Method: Symbolic Information Processing and Self-Organizing Systems in Early Artificial Intelligence Modeling*, in *The Me-*

[8] See Zwart *et al.* 2016.

chanical Mind in History, ed. by P. Husband, O. Holland, M. Wheeler, MIT Press, Cambridge MA, pp. 219-258.
Craik K.J.W. 1962, *The Mechanism of Human Action*, in *The Nature of Psychology,* ed. by S.L. Sherwood, Cambridge University Press, Cambridge.
De Jong K.A. 2006, *Evolutionary Computation: A Unified Approach*, MIT Press, Cambridge MA.
Freemont P., Kitney R. 2012, *Synthetic Biology: A Primer*, Imperial College Press, London.
Hofstadter D.R., FARG 1995, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, Basic Books, New York.
Holland J.H. 1975, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge MA.
Keller E. Fox 2002, *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*, Harvard University Press, Cambridge MA.
Langton C.G. 1989, *Artificial Life*, in *Artificial Life, Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, Sept. 1987,* ed. by C.G. Langton, Addison-Wesley, Redwood City, pp. 1-47.
Mitchell M. 2006, *Complex Systems: Network Thinking*, in «Artificial Intelligence», 170(18), pp. 1194-1212.
Nolfi S., Floreano D. 2000, *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*, MIT Press, Cambridge MA.
Rumelhart D.E., McClelland J.L., PDP Research Group 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge MA.
Trianni V., Tuci E., Passino K.M., Marshall J.A.R 2011, *Swarm Cognition: An Interdisciplinary Approach to the Study of Self-organizing Biological Collectives*, in «Swarm Intelligence», 5, pp. 3-18.
Turing, A.M. 1948, *Intelligent Machinery.* Internal report of National Physics Laboratory, (reprinted in Copeland 2004, pp. 410–432).
Turing A.M. 1952, *The Chemical Basis of Morphogenesis*, in «Philosophical Transactions of the Royal Society B», 237, pp. 37-72.
Von Neumann J. 1951, *The General and Logical Theory of Automata*, in *Cerebral Mechanisms in Behaviour: The Hixon Symposium*, ed. by L.A. Jeffress, Chapman & Hall, London; J. Wiley, New York, pp. 1-13.
Wiener N. 1948, *Cybernetics, or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge MA.
Zwart H., Landeweerd L., Lemmens P. 2016, *Continental Philosophical Perspectives on Life Sciences and Emerging Technologies*, in «Life Sciences, Society and Policy», 12:8.

III

Bodily Boundaries and Beyond: Machines Extend the Bodies and their World?

The Organism, Maker of Machines

Fiorenza Lupi

Machine and Organism is the second of three lectures held by Georges Canguilhem at the *Collège philosophique* in 1946-1947 and then collected in *The Knowledge of Life* in 1952. It opens with the remark that «The problem of the relations between machine and organism has generally been studied only in one direction» (Canguilhem 2008, p. 76): in this analogy, the living organism has always been assimilated to the machine, considering the latter «a simple application of a knowledge conscious of its import and certain of its effects» (*ibidem*). Rather, suggests Canguilhem, it is the construction of a technical object that should be understood starting from the functioning of organisms.

Canguilhem's "U-turn", far from attributing anthropomorphic characteristics to technical objects, aims to shield the organism from biological determinism, which considers it a fully decipherable machine. Thus, the author's operation is not a simple reversal of the relationship, but rather a rethinking that focuses on the specificity of the human being and of the qualities and performances of its sensitivity.

In support of this interpretation, I will consider the text of *Machine and organism* in the framework of other two of Georges Canguilhem's philosophical works on technique: *Descartes et la technique* (1937) and *The Role of Analogies and Models in Biological Discovery* (1963).

The technological problem. Life, science and technique

Canguilhem observes that it is not possible to address the *biological* problem of the organism-machine without approaching the *technological* problem that is the relationship between technique and science. What comes first, science or technique? He recognizes in the technique the expression of human creativity. In his early writings (1926-1939), he had already defined the technical act as a daily form of creation. In *Activité technique et creation*, he wrote that «the technique is the unreflective[1] experience unconsciously directed

[1] Here "irréfléchie", has been literally translated but it means: spontaneous, meaning an experience

towards creation» (Canguilhem 2011a, p. 502, my translation). Far from devaluing the technical invention, this description clarifies its relationship with science: the construction of an instrument is not the mere application of fundamental principles, but manual and practical ingenuity.

Canguilhem criticized the transfer of the Cartesian theory of knowledge into the conception of technical action. In the first lines of *Descartes et la technique,* he wrote: «Is the technical activity a simple extension of objective knowledge – as it has become common to assume after the positivist philosophy, or is it the expression of an original "power", creative after all, and for which science sometimes and only afterwards draws up a research programme or a precautionary code?» (Canguilhem 2011b, p. 490, my translation). Descartes despises art without explanation, inventors without method, and he very much distrusts craftsmen. However, as Canguilhem remarks, in *La Dioptrique* (1637) Descartes shows awareness of a form of relationship between knowledge and technical construction other than that which derives the latter from the former. The starting point of theoretical optics was in fact the invention of the telescope, due to experience and luck. In the transition from theory to practice the possible knowledge (even an ideally perfect knowledge) cannot eliminate all imperfections from technical realization. «Each technical synthesis should include the unpredictable and the unexpected, by operating on parts of the organisms whose deduction may not be integral» (Canguilhem 2011b, p. 496, my translation): the technique here is the model for that creative activity, which can also be found in art and life. Not only does the technical practice precede the scientific discovery, but it is also its necessary condition: the theory, in fact, follows, as an interrogation on the failures and mistakes (creative themselves as knowledge opportunities) of the technical act.

In *Machine and organism*, Canguilhem insists on the need for a technical mediation at the foundation of the scientific knowledge, by mentioning §43 of the *Critique of Judgment*, where Kant defines the peculiarity of human technique:

> What one can do the moment one only knows what is to be done, hence without anything more than sufficient knowledge of the desired result, is not called

not subject to rational activity.

> art[2]. To art that alone belongs for which the possession of the most complete knowledge does not involve one's having then and there the skill to do it. Camper describes very exactly how the best shoe must be made, but he, doubtless, was not able to turn one out himself (Kant 2007, p. 133).

But if this is true, the statement that the machine is «a simple application of a knowledge conscious of its import and certain of its effects» (Canguilhem 2008, p. 76) falls apart. Ingenuity, Canguilhem says, is as inexplicable in its formative process as life. To make use of a machine to explain a biological function is not really to apply a knowledge, and therefore does not entail any certainty about its effects. We may define a technical object «as an artificial construct, a work of human being, whose essential function depends on mechanisms» (*ibidem*).

Canguilhem's second step is to address the relationship between mechanism and finalism: «The mechanical explanation of the functions of life historically presupposes [...] the construction of automatons, whose name signifies at once the miraculous character and the apparent self-sufficiency of a mechanism transforming an energy that is not [...] the effect of a human or animal muscular effort» (*ibidem*, p. 78). However, according to Canguilhem there is no real opposition between mechanism and finalism: in fact, although the functioning of the automaton cannot be reduced to relations of pure causality, its construction would be impossible without a purpose, or without a human being: «A machine cannot replace another machine. The more limited the purpose, the more the margin of tolerance is reduced, and the more hardened and pronounced the purpose appears to be. In the organism, by contrast, one observes [...] a vicariousness of functions, a polyvalence of organs» (*ibidem*, p. 89). Because there is no such thing as a machine able to build other machines, explaining organisms through mechanical models is a tautology, just like explaining an organ using an organ[3].

In *The normal and the pathological*, the author emphasizes the fact that in the case of the living being we cannot talk about normality as conformity to pre-established rules, but we need to introduce the concept of normativity as ability to create new norms replacing the existing ones. "Normal"

---

[2] The German word used by Kant is *Kunst*, which describes any human work, not only the art. We can translate art with τέχνη in general.

[3] It is interesting to note the etymology of the word "organ": from the Latin *organum* which means tool [Lat. *ŏrgănum*, Gr. ὄργανον (closer to ἔργον «work»)].

changes in relation with individual conditions. A good example is that of individuals with only one kidney: organisms that have been able to break the "rule of two kidneys", tolerating (the tolerance: what is missing in the machines) the loss of one of them and inventing the norm that has allowed them to keep living with just one kidney. «Life tolerates monstrosities», whereas «there is no machine monster. There is no mechanical pathology» (*ibidem*, p. 90).

The biological problem. Machines, the living being's organs

In organisms, therefore, phenomena of self-construction, self-regulation, self-preservation and self-repair occur, while all functions of technical objects need human intervention in order to be carried out. However, the living being "needs" the machines too, just in the same way as he needs his organs. In *Machine and Organism*, Canguilhem refers to *Milieu et techniques* by André Leroi-Gourhan, stating that «the last chapters of this work constitute what is today the most striking example of a systematic and duly detailed attempt to bring biology and technology together» (*ibidem*, p. 95). Canguilhem, like Leroi-Gourhan, believes that machines can be considered organs of the human species. Consequently, it is human sensitivity which plays the main role, as it is characterised, since the very beginning, by the ability to extend itself through prostheseis. Like the amoeba, which pushes out of its mass an extension that captures external objects, so human *aisthesis* externalizes itself in technical tools. In both cases the contact by touch is central, «but while the expansion of amoeba always pulls its prey toward the same digestive process, between the matter to be dealt with and the technical thought that envelops it are created, in each circumstance, specific organs of percussion» (*ibidem*, p. 94). The technique was thus born as a «need for the living being» rather than as an application of knowledge:

> The rationalization of techniques makes one forget the irrational origin of machines. And it seems that in this area, as in any other, one must know how to cede a place to the irrational, even and especially when one wants to defend rationalism (*ibidem*, p. 95).

The role of analogy: a normative act

Canguilhem interprets the prevalence in biology of analogical on mathematical models as an evidence that biological entities resist analysis: they are phenomena whose complexity cannot be broken down. The analogical method reduces the organic forms to more familiar mechanisms, borrowing these models from technological experience. What should an analogical model do? Replicate, not only the effects of the organic function, but also the means of its action (the construction of a pattern). A successful model is, for Canguilhem, Vaucanson's flute-player, mentioned by Condorcet:

> The Academy of Science was charged with examining the automaton, and decided that mechanism used to give forth the sounds of the flute rigorously executed the same operations as someone playing a flute, and that the mechanic had imitated at once the effects and the methods of nature (Canguilhem 1963, p. 511).

Condorcet had remarked that «the genius of mechanics consists principally in imagining and disposing in space different mechanisms which must produce a given effect and which serve to regulate, distribute, and direct the motive power» whereas «an artist owes his talents or success to practice» (*ibidem*). For Condorcet, thanks to the design of schemes, it is possible to invent a mechanical masterpiece "in theory"; in the same way as one calculates the movements of a star that one has never seen. Condorcet emancipated technique from execution, while for Canguilhem the opposite is true: the technical invention does not consist in the application of laws, but in a practice that leads to the discovery and creation of rules. As we have seen, this consideration also means that

> By considering technique to be a universal biological phenomenon and no longer only an intellectual operation of a human being, one is led [...] to affirm the creative autonomy[4] of arts and crafts from any knowledge capable of appropriating them so as to apply itself to them or informing them so as to multiply their effects (Canguilhem 2008, p. 96).

[4] From the Greek αὐτονομία "independence", noun of quality from αὐτόνομος "independent, living by one's own laws", (from gr. αὐτός "self" and νόμος "law"). In *The normal and the pathological* it is the ability of the organism to invent his/her own rules of life and build his/her own environment.

In order to explain the functioning of the analogical model, Canguilhem gives the example of a nerve as a fluid current conductor. In the analogy between electricity conduction and blood circulation, the action of the flow is only supposed, not perceived. We need the model when there is no possibility of analytical breakdown. It replaces the representation and makes it possible to consider the blood flow as if[5] it were an electrical flow. A good model must not assume the identity between two objects; there is no «isomorphism of the theories», because the rules of the organism and those of its mechanical model remain valid and verifiable each in its own area. Many parts of the body can be considered *as if* they were machines, but the organism *is not* a machine. Canguilhem quotes Adrian's remark: «What we can learn from the machines is how our brain must differ from them» (Canguilhem 1963, p. 516).

What is the error in which one may incur by using the analogy between machines and organisms? By saying "The organism is a machine" one projects on the living being the aspect of predictability, which seems to characterise the mechanical action. On the other hand, by saying, "The machine is an organism", one thinks of the machine as a completely autonomous form of life. For Canguilhem, even cybernetics is not safe from this danger. Whilst recognizing that some machines are capable of self-organization – as in Rosenblueth, Wiener and Bigelow's study on the anti-aircraft fire control system, interpreted as an example of a machine able to change its behaviour, by simulating the flexibility of the living being –, Canguilhem believes that we cannot underestimate the role of the "as if" in the analogy. If for Rosenblueth, Wiener and Bigelow there is identity between machine and organism in regard to the behaviour, for Canguilhem

> it is worth noting that recent apologists for the heuristic efficiency in biology – especially in neurology – of cybernetic mechanisms and of feedback models, consider the construction of classical automata (that is to say those lacking some sort of feedback mechanism) capable of simulating, within the limits of one or more rigid programmes, animal behaviour or human gestures, to be just a craze, without any scientific interest; merely a pastime (*ibidem*, p. 510).

Moreover, he writes, quoting von Neumann:

[5] Note the proximity to the "als ob" of Kantian philosophy. On this interesting similarity, see Cavazzini 2007.

> the structure of natural machines (organisms) is such that failures of function do not affect general behaviour. Regenerative functions, or, failing this, the supplementing of the insufficiency of one organ by another, compensate for the destruction or the breakdown of certain elements. A lesion of the organism does not necessarily abolish its plasticity. The same cannot be said of machines (*ibidem*, p. 516).

Final remarks

The analogy between organisms and machines may be useful if it is conceived as a *normative* (and so, able to create new rules) act rather than a *normalizing* act, that is as the desire to control and predict the behaviour of the living being. Canguilhem insists on the originality and creativity of the living being as opposed to the machines and recommends caution in the use of the analogy between machine and organism. However, in all his writings on the subject, he reiterates the importance of the technical phenomenon for the human being. The relationship between human beings and technical objects, the features shared by bodies and machines and «the technological structure of human perception» show, once again, «the condition of the human being – an organism, but a maker of machines» (*ibidem*, p. 519).

The relationship between organisms as makers of machines and technical objects will be further addressed by Gilbert Simondon, one of Canguilhem's disciples. According to him, machines have a special "mode of existence" and are able to evolve... but only if they have a human being by their side.

References

Canguilhem G. 1963, *The Role of Analogies and Models in Biological Discovery*, in *Scientific Change,* Symposium on the History of Science University of Oxford, 9-15 July 1961, ed. by A.C. Crombie, Heinemann, London, pp. 507-520.

Canguilhem G. 1991, *The Normal and the Pathological* (1943), Zone Books, New York.

Canguilhem G. 2008, *Machine and Organism* (1952), in G. Canguilhem, *The Knowledge of Life*, Fordham University Press, New York, pp. 75-97.

Canguilhem G. 2011a, *Activité technique et création* (1938), in G. Canguilhem, *Oeuvres complètes,* t. I (*Ecrits philosophiques et politiques 1926-1939*), Vrin, Paris, pp. 499-506.
Canguilhem G. 2011b, *Descartes et la technique* (1937), in G. Canguilhem, *Oeuvres complètes* t. I (*Ecrits philosophiques et politiques 1926-1939*), Vrin, Paris, pp. 490-498.
Cavazzini A. 2007, *Analogia, macchina e organismo,* in G. Canguilhem, *Il vivente, l'analogia, le scienze. Studi di archeologia dei saperi*, Mimesis, Milano, pp. 45-71.
Descartes R. 2005, *Discours de la méthode* (1637), Vrin, Paris.
Kant I. 2007, *Critique of Judgment (*1790), transl. by J.C. Meredith, Oxford University Press, New York.
Leroi-Gourhan A. 2000, *Évolution et techniques,* vol. 2 (*Milieu et techniques*), Albin Michel, Paris.
Rosenblueth A., Wiener N., Bigelow J. 1943, *Behaviour, Purpose and Teleology*, in «Philosophy of Science», 10(1), pp. 18-24.
Simondon G. 1958, *Du mode d'existence des objets techniques*, Aubier, Paris.

Towards a Philosophy of Interaction. The Relation etween Organism, Technique and Environment in Gilbert Simondon

Elisa Binda

The philosophy of Gilbert Simondon (1924-1989) is increasingly appreciated as a valuable contribution to the contemporary philosophical debate about technique. Within the studies in the philosophy of technology, as well as in modern media studies, the reference to Simondon's work is more and more frequent. The focus of this renewed attention to his work mostly concerns his theory about the technical objects and the environment to which they are associated (Simondon 1958). In fact, Simondon's observations devoted to the configuration of human sensitivity could be misunderstood or underestimated if one does not keep in mind the strong connection he establishes between human sensitivity and technique, which is the focus of the present contribution. We believe that the analysis of such an issue is worth of attention nowadays, as human sensitivity is increasingly interconnected with the use of new technologies and is therefore more externalized and interactive than in the past.

Techno-aesthetics and associated environments

The point of departure for this investigation is a short text that Simondon wrote in 1982. It is a letter addressed, but never sent, to Jacques Derrida. In this text, Simondon makes a rather radical statement: «The techno-aesthetic feeling seems to be a category that is more primitive than the aesthetic feeling alone, or than the technical aspect considered from the angle of functionality alone (which is an impoverishing perspective)» (Simondon 2012). The statement that the techno-aesthetic feeling is more original than both the pure aesthetic feeling and the bare technical aspects underscores that human sensitivity, especially if considered according to its species-specific characters, is in principle connected with technical elements, that is, externalized. The kind of aesthetics proposed by Simondon, then, cannot be described as a theory concerning solely the "fine arts", as the reference in the letter to the Greek term *aisthesis* (sensitivity, sensation) highlights. Aesthetics is thus reinstated as a reflection with a larger scope, encompassing experience in general, and investigates the specific way humans interact with the world through their peculiar sensitivity. According to Simondon, this way of interacting with the world is eminently technical.

Among Simondon's essential theoretical references, we must consider the paleoethnologist André Leroi-Gourhan, who strongly influenced Simondon's reflection on technique[1]. This reference helps understanding the statement concerning the original character of techno-aesthetics. In fact, Leroi-Gourhan's research emphasized the reciprocity of hominization and technogenesis. The meaning of human evolution rests for him in the very process of externalization: «the whole of our evolution has been oriented toward placing outside ourselves what in the rest of the animal world is achieved *inside* by species adaptation» (Leroi-Gourhan 1965, p. 235). It is possible to find the specific connotation of human experience and sensitivity in a peculiar relation between "inside" and "outside". According to Bernard Stiegler, whose philosophy is strongly connected with Simondon's reflection,

> Hominization is […] a process of exteriorization which, from the point of view of paleontology, means that the appearance of the human is the appearance of the technical. […]. The movement inherent in this process of exteriorization is paradoxical: Leroi-Gourhan in fact says that it is the tool, that is, *tekhnê*, that invents the human, not the human who invents the technical. Or again: the human invents himself in the technical by inventing the tool – by becoming exteriorized techno-logically. But here the human is the interior: there is no exteriorization that does not point to a movement from interior to exterior. Nevertheless, the interior is invented in this movement; it can therefore not precede it. Interior and exterior are consequently constituted in a movement that invents both one and the other: a moment in which they invent each other respectively […]. The interior and the exterior are the same thing, the inside is the outside, since man (the interior) is essentially defined by the tool (the exterior) (Stiegler 1998, p. 141).

[1] This is confirmed by the reference to André Leroi-Gourhan's works in the bibliography of many of Simondon's texts.

The relation between "inside" and "outside" is, therefore, co-constitutive for the human being. The human processes of externalization have a powerful feedback effect, which reconfigures and modifies human beings.

According to Simondon, this movement of endless reconfiguration between "inside" and "outside" is exactly what constitutes the very process of individuation. His main doctoral dissertation, *L'individuation à la lumière des notions de forme et information*, begins with the assertion that the individual is the result of an endless process of individuation. According to Simondon, the process of individuation moves from a pre-individual natural asset, understood as a generating and creative power, close to that of the Pre-Socratic *physis*. The appearance of the individual is however the result of a coupling with the associated environment. By embedding the individual into the pre-individual, the former appears as a problematic being, which attempts to "solve" itself by individuating, i.e. always configuring individuations anew, thanks to the relation to an associated environment. This environment can be considered as the field from which the individual derives potentialities.

The relation between the individual and the environment is thus fundamentally co-constitutive, since "inside" and "outside" configure each other: «the state of a living being consists in a problem to be solved, of which the individual itself is the solution on the basis of consecutive *montages* of structures and functions» (Simondon 1964, p. 223). Moreover, he adds, «the ontogenetic development itself can be considered as a mediation» (Simondon 1992, p. 317). The concept of techno-aesthetics shows how far, in the course of the process of the human individual's externalization, consecutive *montages* of structures and functions are established through the technical mediation with an environment: «it is for the very fact that the living being is an individual being, embedding its associated environment, that the living being is able to invent; this agency of self-conditioning is at the basis of its capacity of producing objects that are in turn self-conditioning» (Simondon 1958, p. 58). What binds together human beings, technique and environment can be defined as a "transductive relationship".

The concept of transduction is borrowed from physics. With this term Simondon describes the relationship that binds together two or more elements at the very moment of their constitution – the relationship being, however, a necessary condition for this constitution. Therefore, the relation among human beings, technique and environment is profoundly interactive and transformative for each of these three interrelated factors.

Simondon's work *Du mode d'existence des objets techniques* (1958) allows to appreciate how this framework holds also for the techno-aesthetic sensitivity. In this text, Simondon states that a technical object can be considered as such only if it is endowed with a margin of indetermination, which allows it to «be sensitive to the information from outside». This open character enables the technical object's interaction with the environment. What Simondon names "associated environment" is constituted within this relation of interaction, which he calls "concretization":

> The adaptation-concretization process is one which causes the birth of an environment, rather than being the result of an already established environment. […] The invention happens because a jump is made and is justified by the relationship which is instituted within the environment it creates. […] It could be said that concretizing invention brings into being a techno-geographic environment, which is the condition upon which the possible functioning of the technical object depends. Therefore the technical object is the condition of itself as condition for the existence of this mixed environment that is at once technical and geographical. […] This environment, both technical and geographical, can be named "associated environment" (*ibidem*, p. 55).

The constitution of associated environments is among the most remarkable consequences of all human technical externalizations, and it is based on the fact that human sensitivity is inherently a techno-aesthetic sensitivity. The individual reconfigures its environment through the technical object, but she is in turn implied in this reconfiguration. According to Simondon, «technology makes a relational function appear». Such a function concerns what he calls the relation of "couplage" that holds together «organism and environment, human being and world» (Simondon 2005, p. 84).

Culture et technique: the feedback between inside and outside

In a minor text, *Culture et technique* (1965), Simondon further emphasizes the issue of the transductive relation, in order to show how far the exclusion of technique from the domain of culture is incorrect, by analyzing what he calls «feedback effects».

Here Simondon proposes an important definition of "technique". "Technique" is when «man acts on the environment he is exploiting, transforming, organizing; in this case, he doesn't act on himself that through that *charge* which is the environment » (Simondon 2014, p. 317). Through the modifications imposed to the environment by technical mediations, human beings are able to act on themselves, becoming engaged in a process of individuation: a feedback is produced. As a consequence, technique cannot be conceived just as a

> means, but rather as an act, a phase of an activity of relation between a human being and its environment. During this phase, the human being stimulates its environment by modifying it. Such a modification develops further, and the environment so modified offers the human being a new field for action, demanding a new adaptation [...]. Making its way through the environment, the energy of the technical gesture feeds back into the human being and allows him to change, to evolve (*ibidem*, p. 320).

For Simondon, such feedback effects, produced through an externalization of the technical gesture, are effective on the human cognitive and sensitive experience: «everything happens as if the bodily scheme of the human species had been modified, expanded, as if it had got new dimensions. The level of magnitude changes; the perceptual pattern enlarges and differentiates. New schemes of intelligibility arise […]. It is embodiment» (*ibidem*, pp. 324-325). The feedback action realized by the technical gesture produces then «new schemes of intelligibility», i.e. a significant expansion of our cognitive performances.

This expansion is followed, however, by an equally significant reconfiguration of the human perceptual agency. As a result, human sensitivity is modified. The human bodily scheme expands and is enriched by new dimensions. Human «perceptual pattern» expands and increases its differentiations. These modifications, embodied into the individual, favor the emergence of new approaches both to the cognitive and sensorimotor environment. In this way, an important aspect of human evolution in its historical course emerges, that it occurs «through the technical gesture». Human and technological evolution are related by a mutual interaction. Such a mutual configuration entails also the modification, the expansion – sometimes, even the reduction – of our perceptual and sensitive asset, which is constantly reorganized within the technical environment to which it happens to be historically merged.

Simondon's reflections on the human interactive relationship with the technique are crucial for the contemporary age, in which human sensitivity is increasingly externalized and open. The relationship between "inside" and "outside" is in fact becoming more and more transductive within modern medial environments.

Furthermore, several of Simondon's issues appear to echo some aspects of the Extended Mind Theory (EMT). While defining the human being as a «Natural Born Cyborg», Andy Clark – one of the most influential supporters of EMT – has not in mind humanoids empowered by mechanical devices and programmed brains. Rather, the very meaning of EMT emphasizes that human relationship to technique is eminently constitutive, species-specific, because of a naturally technically altered, i.e. techno-aesthetic, sensitivity. Clark refers to feedback processes that are available in the permeability between "inside" and "outside" environment and comprehend all human technical mediations:

> we discern two distinct, but deeply interanimated, ways in which biological cognition leans on cultural and environmental structures. One way involves a developmental loop, in which exposure to external symbols adds something to the brain's own inner toolkit. The other involves a persisting loop, in which ongoing neural activity becomes geared to the presence of specific external tools and media (Clark 2003, p. 78).

Clark insists on the strongly affective dimension of the feedback effects engendered by the way in which human beings interact with technologies, which is characterized by «some kind of local, circular process in which neural commands, motor actions, and sensory feedback are closely and continuously correlated» (*ibidem*, p. 104). This makes possible to shape «the complex feedback loops that connect action-commands, bodily motions, environmental effects, and multisensory perceptual inputs. It is the two-way flow of influence between brain, body, and world that matters, and on the basis of which we construct (and constantly re-reconstruct) our sense of self, potential, and presence. The biological skin-bag has no special significance here. It is the flow that counts» (*ibidem*, p. 114).

Both Simondon and Clark, then, believe that relationship, interaction, is the pivot of these feedback effects. Technical mediations reorganize the human experience of the world.

References

Clark A. 2003, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, Oxford University Press, New York-London.
Leroi-Gourhan A. 1993, *Gesture and Speech,* MIT Press, Cambridge-.
Simondon G. 1958, *Du mode d'existence des objets techniques*, Aubier, Paris.
Simondon G. 1992, *The Genesis of the Individual,* in G. Simondon, *Incorporations,* ed. by J. Crary, S. Kwinter, Zone Books, New York, pp. 297-319.
Simondon G. 2005, *L'invention dans les techniques,* Éditions du Seuil, Paris.
Simondon G. 2012, *On techno-aesthetics*, in «Parrhesia», 14, pp. 1-8.
Simondon G. 2014, *Culture et technique* (1965), in G. Simondon, *Sur la technique,* Puf, Paris, pp. 315-329.
Stiegler B. 1998, *Technics and Time,* vol. 1: *The Fault of Epimetheus,* Stanford University Press, Stanford, Calif.

Mixed Systems and Interplay. Norbert Wiener meets Walter Benjamin

Francesco Restuccia

Norbert Wiener, the American mathematician (1894-1964), and Walter Benjamin, the German philosopher (1892-1940), never met and probably never heard of each other. However, their thoughts about the interaction between human beings and machines have a few interesting similarities. They are both concerned with the impact of technology on society[1] and they consider the second industrial revolution the turning point in the evolution of the relationship between human and machine.

However, they followed different routes and had different goals. Wiener approached technology from the perspective of control engineering theory and only after World War II started questioning it from an ethical point of view. Benjamin never studied any particular machine and only approached technology from an "anthropological" perspective, considering especially the way human beings deal with their own production, including crafts and art (*techne*). The viewpoints Wiener and Benjamin share are not many; yet they are significant for anyone wishing to work on these issues in an interdisciplinary way.

The turning point

In *The Human Use of Human Beings*, his first book about the ethical and sociological implications of cybernetics and technology, Norbert Wiener distinguishes «the older machines, and in particular the older attempts to produce automata», from the «modern automatic machines such as the controlled missile» (Wiener 1950, p. 22). The former functioned «on a closed clockwork basis» and did not have any interaction with the environment. The latter possess sense organs, which enable them to receive messages from, and interact with, the environment. If the engine is the essential element of the

[1] The social implications of cybernetics are evident also in its etymology: the term, that Wiener himself proposed, is based on the Greek word *kybernétes* (steersman, captain), which is related to the Latin word *gubernum* (government).

first industrial revolution, substituting the labor of slaves and animals with the energy of the machine, the second industrial revolution can find its icon in the photoelectric cell (*ibidem*, p. 23). A few pages further Wiener describes these new machines through two general features:

> One is that they are machines to perform some definite task or tasks, and therefore must possess effector organs (analogous to arms and legs in human beings) with which such tasks can be performed. The second point is that they must be *en rapport* with the outer world by sense organs, such as photoelectric cells and thermometers, which not only tell them what the existing circumstances are, but enable them to record the performance or nonperformance of their own tasks. This last function, as we have seen, is called *feedback*, the property of being able to adjust future conduct by past performance (*ibidem*, pp. 32-33).

The automata – the older machines – only execute what they are programmed for and they need humans to regularly adjust their functioning. On the contrary, machines provided with self-regulation systems – modern machines – are not only more efficient, but more autonomous: one can now be surprised by the performances of a machine such as a chess-player computer.

A similar distinction between a dependent and an autonomous technology can be found in Walter Benjamin's well-known essay *The Work of Art in the Age of Its Technological Reproducibility* (1936). This essay is much more than a reflection on art; Benjamin thinks anew about the way technology transforms our experience of the world. Even though human life has always been somehow technical, two kinds of technology can be recognized, according to the sort of interaction they establish with human beings: an older one based on mastery over nature, and a second one based on interplay.

> Whereas the former made the maximum possible use of human beings, the latter reduces their use to the minimum. The achievements of the first technology might be said to culminate in human sacrifice; those of the second, in the remote-controlled aircraft which needs no human crew. The results of the first technology are valid once and for all (it deals with irreparable lapse or sacrificial death, which holds good for eternity). The results of the second are wholly provisional (it operates by means of experiments and endlessly varied test procedures) (Benjamin 2008, p. 26).

The *first technology* originated in ancient times, but far from being restricted to the past, it is still present today, every time something is accomplished «once and for all». The *second technology*, on the contrary, is quite recent, because it needs receptors to function by itself, reducing the use of human beings to the minimum. Obviously, a remote-controlled aircraft is not yet completely autonomous, since it needs a ground control, but it is considered by Benjamin a first step in this direction.

It is remarkable that Wiener and Benjamin employ the same vocabulary in addressing this issue: they both are interested in the new machines not for their efficiency, but because they make a «human use of human beings» possible (Wiener 1950), which means reducing «their use to the minimum» (Benjamin 2008, p. 26). The aim of first technology is to transform nature, while the second technology aims at functioning within the world: the former tries to adapt nature to itself, the latter tries to adapt itself to the world.

The second technology operates by «endlessly varied test procedures», in an experimental way. Benjamin considers tests a distinctive feature of the way of living of our society, in sport, in acting performances, and in the work process that, «especially since it has been standardized by the assembly line, daily generates countless mechanized tests» (*ibidem*, p. 30). Test performances are based on a process similar to a feedback effect[2]: the behaviour is periodically compared with the result to be achieved, and the success or failure of this result changes the behaviour of the performer. This is why the results of second technology «are wholly provisional». According to Benjamin tests confer to any act a playful dimension. «The origin of the second technology lies at the point where, by an unconscious ruse, human beings first began to distance themselves from nature. It lies, in other words, in play» (*ibidem*, p. 26).

Interplay

One *uses* the older machines, but one *plays* with the new ones. Interacting with the new machines has a recreational aspect that is not present in the clockwork-like machines, since these latter are foreseeable. The behaviour of the apparatus does not depend entirely on our inputs, but also on its inacces-

[2] Baudrillard 1993 compares Benjamin's concept of test performance to feedback, but unlike Wiener, he considers a feedback-based society a non-democratic one.

sible internal program, and especially on the environment, and it is therefore partly unpredictable for the user. This potential surprise, or, as Benjamin calls it, «the shock effect», creates an emotional expectation and induces «heightened attention» (*ibidem*, p. 53): the one who interacts with an apparatus is both alert, since he is expecting a partly unpredictable result, and distracted, *zerstreut*, which in German also means entertained. The kind of interaction Benjamin is thinking of is not an intellectual one: he writes about a «physical shock effect» and «primarily tactile» distracting elements (*ibidem*, p. 39).

The playful aspect of the second technology is not restricted to entertainment, but it also includes learning, just as children's games have both a recreational and an educational dimension. Playing with the new machines, the user improves his «know-how» (Wiener 1950, p. 183): the interaction with the apparatus is a «true training ground» (Benjamin 2008, p. 41). But what do we learn? What do we need to be trained for? Of course, we need to learn how to handle the machines themselves. However, one needs to be trained not only to use the machines properly, but especially «to preserve one's humanity in the face of the apparatus […], for the majority of city dwellers, throughout the workday in offices and factories, have to relinquish their humanity in the face of an apparatus» (*ibidem*, p. 31). The loss of humanity, according to both our thinkers, is due to the lack of *responsibility*, which has to be understood in the sense of *capability to respond*.

> I have spoken of machines, but not only of machines having brains of brass and thews of iron. When human atoms are knit into an organization in which they are used, not in their full right as responsible human beings, but as cogs and levers and rods, it matters little that their raw material is flesh and blood. *What is used as an element in a machine, is in fact an element in the machine* (Wiener 1950, p. 185).

Wiener, just like Benjamin, thinks that only a machine, which is able to adapt itself to its environment, can establish with the user a "human" interaction, but the user also needs to learn how to dialogue with it. The human being is testing the apparatus, while the apparatus is testing the human performance: they are both learning from each other.

The main example of a learning apparatus, in *God & Golem Inc.* (Wiener 1964), is a computer that was developed by A. L. Samuel of IBM Corporation in 1959, and that could play checkers. The computer, just like the human player, improves its performances by its own experience of the actions of the other player (Figure 1). In this case, it is apparent that the human is not using the machine, as the computer is not using the human being. *Use* is a term that belongs to the first technology, while at this stage we should rather talk of interplay. «The first technology really sought to master nature, whereas the second aims rather at an interplay between nature and humanity» (Benjamin 2008, p. 26). *Zwischenspiel* in German means interplay, ludic interaction, but it is also employed to mean an *intermezzo*, a musical interlude that separates two parts and at the same time relates them.

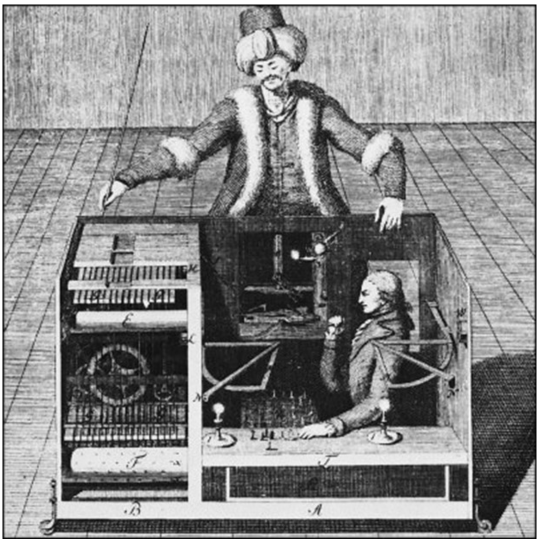


Figure 1. The Mechanical Turk, Von Kempelen's fake Automaton Chess Player, from Racknitz 1789, quoted by Benjamin 1968, p. 253.

Mixed systems

The relationship between human and machine establishes a new unity that includes the two components and the environment: the camera is connected to the photographer and the photographed subject, the car to the driver and the street. Modern machines, according to Wiener, are «systems of a mixed nature, involving both human and mechanical parts» (Wiener 1964, p. 76).

Technical apparatuses connected to an organism are usually called prostheses. Wiener distinguishes them into three kinds. A simple, mechanical substitution of a missing limb, such as a wooden leg, is the most trivial case. A more interesting one is the prosthesis that substitutes for muscles and damaged sense organs, such as a robot hand connected to the nervous system. But the third example is the most important: «this type of engineering need not to be confined to the replacement of parts that we have lost. There is a prosthesis of parts which we do not have and which we never had» (*ibidem*). On our airplanes, we have the wings of an eagle, thanks to our sonars we navigate like dolphins. This enhancement is not only for the single individual, but also for groups of people and for the whole society.

In a similar way Benjamin writes that a «new, historically unique collective» is born, «which has its organs in the new technology» (Benjamin 2008, p. 45). To refer to the connection between this new collective and its technological organs, its prostheses of parts which it never had, Benjamin uses the term *innervation* that he borrows from Freud's early writings. It means both the distribution of nerves in an animal to any of its parts and the act of stimulating an activity in any of its organs. This deep connection is still more a project than a reality, and that is why Benjamin writes about «efforts at innervation»: a stimulation that expects a response – a playful training again. «Just as a child who has learned to grasp stretches out its hand for the moon as it would for a ball, so humanity, in its efforts at innervation, sets its sights as much on currently utopian goals as on within reach» (*ibidem*). A seemingly useless gesture like stretching out one's hand for the moon may actually reveal itself as a training that will eventually help learning how to better grasp a ball, but at the same time it reveals that one could grasp much more than a ball.

> Dealing with apparatus also teaches them that technology will release them from their enslavement to the powers of the apparatus only when humanity's whole constitution has adapted itself to the new productive forces which the second technology has set free (*ibidem*, pp. 26-27).

> Because this technology aims at liberating human beings from drudgery, the individual suddenly sees his scope for play, his field of action, immeasurably expanded. He does not yet know his way around this space. But already he registers his demands on it (*ibidem*, p. 45).

Benjamin offers two examples of second technology: the remote-controlled aircraft and the movie camera. They both require interplay and establish together with the human being a mixed system: they expand the human field of action (*Spielraum*: a space for playing) as prosthesis of parts humans never had. Thanks to his new mechanical eye the human being can now extend movements with slow motion, and expand space with enlargement, disclosing his «optical unconscious» (*ibidem*, p. 37).

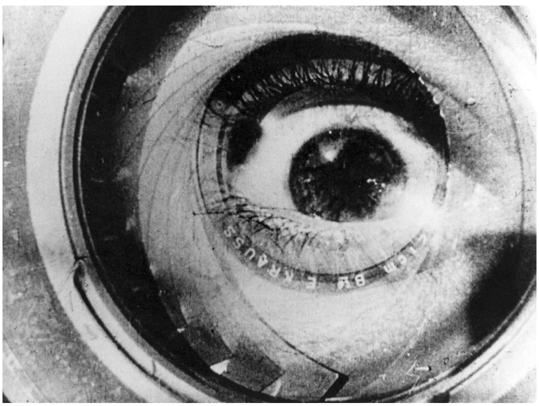


Figure 2. Frame of Dziga Vertov's Man with a Movie Camera, 1929, a director quoted by BENJAMIN 1968, p. 231.

Know-how and know-what

The more technology is automatized, the more our field of action is extended. Do we still need to feel responsible for a completely automatized technology? Does it «take over from us our need for difficult thinking»? Only if we believe that it thinks *for* us and not *with* us, we will make this mistake. Automation is «literal-minded»: a modern apparatus will reach its goal with unforeseen strategies, but it will only reach *that* goal. «A goal-seeking mechanism will not necessarily seek *our* goals, unless we design it for that purpose» (Wiener 1964, p. 63). The unpredictable results of such a machine are very interesting, because they show us something we did not think of; however, this can also be very dangerous. That is why the programming of an apparatus is a very important task.

> If you're playing a war game with a certain conventional interpretation of victory, victory will be the goal at any cost, even that of the extermination of your own side, unless the condition of survival is explicitly contained in the definition of victory according to which you program the machine (*ibidem*, p. 60).

Automation should not be a way of delegating our concerns to machines; on the contrary, we should learn to use our new extended field of action to face these concerns in a new playful way together with the apparatus. «Vital questions affecting the individual – questions of love and death which had been buried by the first technology – once again press for solutions» (Benjamin 2008, p. 45). What we should try to understand in our interplay with the second technology is what we want, no matter if it is or it is not within reach.

> Our papers have been making a great deal of American "know-how" ever since we had the misfortune to discover the atomic bomb. There is one quality more important than "know-how" and we cannot accuse the United States of any undue amount of it. This is "know-what" by which we determine not only how to accomplish our purposes, but what our purposes are to be (Wiener 1950, p. 183).

References

Baudrillard J. 1993, *Symbolic Exchange and Death*, SAGE Publications, London.
Benjamin W. 1968, *Illuminations*, Schocken Books, New York.
Benjamin W. 2008, *The Work of Art in the Age of Its Technological Reproducibility*, Harvard University Press, Cambridge MA-London.

Cordeschi R. 2002, *The Discovery of the Artificial*, Kluwer Academic Publishers, Dordrecht.
Duttlinger C. 2007, *Between Contemplation and Distraction: Configurations of Attention in Walter Benjamin*, in «German Studies Review», XXX(1), pp. 33-54.
Montagnini L. 2005, *Le Armonie del Disordine. Norbert Wiener matematico-filosofo del Novecento*, Istituto Veneto di Scienze, Lettere ed Arti, Venezia.
Montagnini L. 2013, *Interdisciplinary Issues in Early Cybernetics*, in *New Perspectives on the History of Cognitive Science*, ed. by C. Pléh, L. Gurova, L. Ropolyi, Akadémiai Kiadň, Budapest, pp. 81-89.
Racknitz J.F. zu 1789, *Über den Schachspieler des Herrn von Kempelen und dessen Nachbildung*, J. G. I. Breitkopf, Leipzig-Dresden.
Rosenblueth A., Wiener N., Bigelow J. 1943, *Behaviour, Purpose and Teleology*, in «Philosophy of Science», X(1), pp. 18-24.
Wiener N. 1948, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge MA.
Wiener N. 1950, *The Human Use of Human Beings*, Houghton Mifflin Company, Boston.
Wiener N. 1964, *God & Golem Inc.: A Comment on Certain Points where Cybernetics Impinges on Religion*, The Riverside Press, Cambridge MA.
Winner L. 1977, *Autonomous Technology*, MIT Press, Cambridge MA.

Only Connect. The Contribution of Michael Tomasello to the Machine-Organism Symbiosis

Francesco De Bei

In his classic 1960 paper, L. Licklider argued the importance of a "symbiosis" man-computer in order to increase the human intellect, thus freeing it from tedious tasks. Today, after more than fifty years, we can count on powerful personal computers, huge online databases and, of course, a ubiquitous Internet. Undoubtedly, the power of our intellect has been increased. It is, however, equally true that, today, computers and technology have ceased being simple tools that can lift us from demanding tasks, and are entrenched in a much more radical way in everyday life. That is, they have become an integral part of the environment.

Over the same period of time it was not only the concept of computer to undergo a transformation, but also the concept of the body has been radically changed. Both clinical and developmental research (e.g., Beebe, Lachmann 2002; Siegel 2012), and cognitive science, especially that ensemble of procedures and research programs that go under the name of *Embodied Cognition* (Shapiro 2011), have taught us to review the boundary between inside and outside, to conceive of a situated mind, simultaneously composed of internal and external processes. In other words, assuming a mind without an environment equates to investigating an empty shell: it is what takes place *between* individuals and *in the* environment that becomes the matter of the mental register. That same matter which, in turn, the mind rearranges and transforms, influencing the exchange with the outside; which, once again, becomes a form of internal information, and so on.

It is clear that this profound change has made the idea of a symbiosis between machine and organism something very different – and probably more fundamental – from the one advocated by Licklider. It is therefore legitimate to ask: how technology and technological objects, increasingly present in the environment and in social practices, can be thought of as an extension of the body and of human cognition?

The advantage of the evolutionary approach to the study of the "extended" quality of human cognition is the ability to shift the question from *as* (the thought is formed) to *how* (it has evolved). This allows locating the point of inquiry in the interaction between the individual and the environment. The mind, as we study it today, has evolved – this the thesis of Tomasello (2008, 2014, 2016) – *first* as an adaptation to a physical environment and *then* to a social environment. In other words, the unique quality of human cognition involves not only an epistemic problem, but (mainly) an ontological one.

From the individual to the shared intentionality

The question from which Tomasello moves can be formulated in this way: what is unique about human thought? His answer lies in the concept of *shared intentionality*: in the entire animal kingdom, only humans are capable of building a network of connections and cooperative motivations with other members of the group. This is due to their ability to empathize; a capacity that, together with the disposition to establish a bond of attachment, reflects an innate predisposition to coordinate thought and behaviour on a cooperative and mutual basis (Bowlby 1969; Tomasello 2009). But to understand how this capability has led to the development of a mind, it is necessary to start from the kind of cognition present in our non-human ancestors (Tomasello 2014).

For a body to adapt to an unpredictable environment, cognitive and decision-making processes are required that enable it to recognize new situations and to adapt to unpredictable events. Faced with a new situation, it must be able first to identify the causal and/or the underlying intentional relations; only from this understanding can then arise an appropriate behavioural response. A cognitively competent body therefore operates according to the classic belief-desire model of the rational action: it must recognize situations that are causally relevant with respect to its aims and values and select appropriate actions to satisfy them.

This flexible and self-regulated cognitive functioning, characteristic of all great apes and our pre-human ancestors, is what Tomasello (2014) calls *individual intentionality*. According to this model, we can say that there is thought when an organism tries to solve a problem (to achieve a goal) not acting directly, but by imagining what would happen in the given situation if it acted in a certain way. In order to possess an individual intentionality an organism must be able (Tomasello 2014): 1) to cognitively represent off-line potential perceptual experiences; 2) to simulate and make inferences, transforming these representations on a causal/intentional/logical level; and 3) to self-monitor these simulated experiences and to assess their specific behavioural outcomes, so as to choose the most appropriate one. It is at this stage that selective pressure acts, because the success or failure of a specific behavioural decision indirectly exposes the underlying processes of representation, simulation and self-monitoring to the test of natural selection.

It should be remarked that, given the action of selective pressure, the content of these cognitive representations is inevitably linked to the external environment. If goals and values are represented as desired situations, then the organism must turn its attention to the perceived environment towards situations relevant to those goals and values. The sight of a banana tree and the absence of predators, for example, represents for a chimpanzee in search of food a *relevant situation* to decide what to do. Given that for an individual intentionality relevance is represented by all those situations that are opportunities or obstacles to the pursuit of the aims of the organism, the determination of relevant situations calls into question the whole way of life of an organism[1] (Tomasello 2014).

This form of rationality is the kind of cognition needed by organisms, such as apes, with competitive interactions with their environment. The human way of life, on the contrary, is organized in a much more cooperative way. According to the hypothesis of Tomasello, it was the selective pressure exerted by these forms of cooperative sociability that turned individual intentionality and cognition in *shared cognition and intentionality*.

This key step occurred thousands of years ago, when an ecological change that reduced the possibility to obtain food has forced the first human beings to choose between procuring food together or go hungry. This new form of interdependence has led to extend the sense of empathy beyond the circle of relatives and friends, to include the companions of collaboration, and brought about new forms of cooperation that required, for their coordination, new forms of cooperative communication which, in turn, produced new types of cognitive representation, inference and self-monitoring – i.e., a new way of thinking.

[1] This idea is directly linked to the Gibsonian notion of *affordances*, but is much broader, as it includes not only the opportunities for action offered directly from the environment, but also many situations that are relevant to the organism in a more indirect way. In the case of the chimpanzee, for example, the absence of a predator is an indirect affordance.

Unlike individual intentionality, shared intentionality is defined by aims and values of *both* participants in the interaction – for example, the will to hunt together – and by their mutual knowledge or "common ground"; that is, to know that that is what you both want. The cognitive model characterized by shared intentionality has therefore a two-level structure (conjunction and individuality). In a collaborative activity, each individual is both the "we" that pursues a joint aim and, at the same time, the individual who has an own role and perspective (Tomasello 2014). In other words, in intentionally shared activity the individual understands what his/her are doing from an "overall overview", not looking at the other from the outside; rather, on one side, every individual represents himself the role and the perspective of the other, and on the other side, he imagines how the other is imagining its role and its perspective[2].

As pointed out by Tomasello (2014), this change has led to the need for new forms of cooperative communications, which have elicited new forms of thought, driven by the "socialization" of all three components of the thinking process: representation, inference and self-monitoring. Regarding representation, the decisive innovation is that both participants in the communicative interaction have to take into account each other's perspective on the situation. Therefore, the act of communication represents prospectively the scene for the recipient. Inferences in collaborative communications have become socially recursive: individuals make inferences about other's intentions regarding their intentional states. Finally, the socialization of self-monitoring is achieved by imagining oneself in the role of the recipient who tries to understand, in order to check whether the communicative act is formulated correctly and is likely to be understood.

Briefly, the transition from an individual to a shared intentionality leads to the development of a cognitive model oriented to the group. Moreover, and more importantly, the cognitive representation, inference and self-monitoring at the basis of this mode of thinking is not based (primarily) on the physical environment, but in established practices of group collaboration. They refer to a social ontology.

[2] It should be noted that at the level of the proximal (psychological) mechanisms: a) the awareness of interdependence does not need to have any role in the decision making process of the individual; the proximal motivation may simply be to cooperate with anyone who has certain characteristics or is in a certain context; and, b) the co-operative act is not motivated by any previous specific act, but only in order to preserve the relationship.

In summary, the concepts of individual and of shared intentionality possess characteristics that naturally provide the cognition with an "extended" quality. In a cognition based on individual intentionality, thinking takes place within the relationship between the individual (the body) and the physical environment of space and quantity (Tomasello, 2014). Analogical inferences such as bigger/smaller, higher/lower, more/less are based on this kind of cognition (e.g., Lakoff, Johnson 1999). In the framework of shared intentionality, inferences are based not on the relationship among physical entities, but between an entity and a wider process or event in which it has a role (Markman, Stillwell 2001). Furthermore, the empathetic base on which rests this form of thought is able to create a "connection" between the members of a group that extends beyond the physical presence. Several clinical data show the impact that the exclusion from the group has on the individual's mental functioning. It is well known, for example, that the perceived loss of belonging to their group – namely a break in the relationship of interdependence – has a negative impact on the individual's ability to self-regulate psychologically and physically (e.g., Cacioppo, Patrick 2009). In a broad sense, what these data suggest is consistent with the assumptions of Tomasello: that is, a break in cooperation – a "disconnection" from the group – can in itself produce severe perturbations in individual mental functioning.

Only connect

The increasingly widespread dissemination of constantly "connected" technological objects – a "network" of information and people – has transformed today's machines from something able to perform demanding or boring tasks to a medium capable of supporting and boosting just those abilities to connection that Tomasello detects at the basis of human cognition. If technological objects based on a "in network" functioning – that is, connected to other systems or to Internet – appear much closer to the analogy of mental functioning, on the other side it is exactly this feature that makes them much more entrenched and in "symbiosis" with individuals. And this in two ways.

Firstly, the amount of information we can access today is vastly larger than in the past. Internet, as well as the so-called *Internet of Things*, have greatly increased the number of operations that can be left to technological objects and the control that we can have on the environment. This change, according to Tomasello, directly involves cognition. In fact, the exponential increase of information and control over the physical objects represents, for the "individual intentionality", part of a decision-making process devoted to the calculation of opportunities and obstacles relevant to the behavioural decision. In other word, what has increased significantly is the amount of direct and indirect *affordances*.

Secondly, the new technological objects provide humans with the ability to communicate constantly with a growing number of individuals and virtual communities (from Facebook to WhatsApp chat), enhancing sharing and collaborative problem-solving. This affects directly the cognitive abilities at the base of a shared intentionality. The result is that the number of recipients (individual and group) on which the inference, representation and self-monitoring is constantly applied has grown dramatically, leading to an increased number of "common land" (from the simple chat to a wider virtual community) to which individuals are constantly "connected."

The machine-body symbiosis today must be thought of as more profound and radical than it was conceivable fifty years ago. The dramatic depth attained by this new machine could produce, along the lines of Tomasello approach, not just an increase in the externalization of cognitive processes, but also a change in the functioning of the mind. In fact, this new form of interaction among individual, technology and environment affects directly the inferences and representations processes underlying human cognition. The issue at stake is evidently to find a new balance in the growing machine-body symbiosis, by developing new forms of adaptation in an environment increasingly saturated of technological objects, and, on a shorter term, to address the basic need to maintain a private and independent space.

References

Beebe B., Lachmann F.M. 2002, *Infant Research and Adult Treatment*, Analytic Press, Hillsdale, NJ.

Bowlby J. 1969, *Attachment and Loss*, vol. 1*: Attachment*, Basic Book, New York.

Cacioppo J.T., Patrick W. 2009, *Loneliness: Human Nature and the Need for Social Connection,* Norton, New York.

Lakoff G., Johnson M. 1999, *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, Basic Books, New York.

Licklider J.R. 1960, *Man-Computer Symbiosis*, in «IRE Transactions on Human Factors in Electronics», 1, pp. 4-11.
Markman A., Stillwell H. 2001, *Role-governed Categories*, in «Journal of Experimental and Theoretical Artificial Intelligence», 13, pp. 329-358.
Shapiro L. 2011, *Embodied Cognition*, Routledge, Oxon, UK.
Siegel D. 2012, *The Developing Mind. Second Edition*, Guilford Press, New York.
Tomasello M. 2009, *Why we Cooperate*, MIT Press, London.
Tomasello M. 2014, *A Natural History of Human Thinking*, Harvard University Press, Cambridge.
Tomasello M. 2016, *A Natural History of Human Morality*, Harvard University Press, Cambridge.

Abstracts

Francesco Bianchini
Research Assistant Professor in Logic and Philosophy of Science, Dept. Philosophy and Communication Studies, University of Bologna, Italy.
Email: francesco.bianchini5@unibo.it

The Synthetic Approach and the Evolution of Cognition

It is usually emphasized that Artificial intelligence (AI) and cognitive science are two distinct fields, which experienced different trends in different times. Older trends were symbolic, computational and representation-oriented, while newer trends are rather sub-symbolic, biologically inspired, dynamical, brain-grounded and environment-oriented, consistent with the embodied cognition paradigm. However, I claim in this paper that the assumption of a biological conceptual framework in the new AI and cognitive sciences is not superimposable with the adoption of evolutionary notions and methods. I further challenge the thesis that the synthetic method represents a unifying framework for AI and cognitive science, especially as far as the evolutionary approaches to the explanation of cognitive capabilities and intelligent behaviour are concerned.

Elisa Binda
PhD in Philosophy, Dept. of Philosophy, Sapienza University of Rome, Italy.
Email: elisa.binda@uniroma1.it

Towards a Philosophy of Interaction. The Relation between Organism, Technique and Environment in Gilbert Simondon

In recent years, the work of Gilbert Simondon has been rediscovered and increasingly appreciated as a valuable contribution to the contemporary philosophical debate about technique. I move in particular from two texts, *Du mode d'existence des objets techniques* (1958) and *Sur la technique* (2014), in order to highlight the characteristic relationship of mutual implication that exists for Simondon among technical object, organism and environment.

Giuseppe Boccignone
Associate Professor, Perceptual Computing and Human Sensing Lab Director, Dept. Computer Science, University of Milan, Italy.
Email: giuseppe.boccignone@unimi.it

Take Another Little Piece of my Heart: a Note on Bridging Cognition and Emotions

Science urges philosophy to be more empirical and philosophy urges science to be more reflective. This dynamics characterized the process of the "discovery of the artificial": in the early days of Cybernetics and Artificial Intelligence (AI) researchers aimed at making machines more cognizant while setting up a framework to better understand human intelligence.
By and large, those genuine goals still hold today, as AI has become more concerned with specific aspects of intelligence, such as (machine) learning, reasoning, vision, and action. However, the field suffers actually from a chasm between two formerly integrated aspects. One is the engineering endeavour devoted to the development of applications. The other is the philosophical debate addressing questions concerning the nature of intelligence. Bridging these two levels can be crucial for a deeper understanding of minds.
An opportunity might be offered by the cogent theme of emotions. Traditionally, computer science, psychological and philosophical research restricted themselves to the investigation of mental processes that do not involve mood, emotions and feelings, in spite of Simon's early caveat that a general theory of cognition must incorporate the influences of emotion. Recent neurobiological findings and technological advances show that the time is ripe to seriously address this promising, albeit controversial, opportunity.

Nicole Dalia Cilia
PhD in Philosophy, Dept. of Philosophy, Sapienza University of Rome, Italy.
Email: nicole.cilia@uniroma1.it

Understand Me or Duplicate Me? Levels of Explanation in Artificial Methodology

The development of new technologies has caused a proliferation of the use of "machines" in various fields. A well-programmed machine is able, if not to replace humans, to assist them in many tasks and to explain many cognitive processes. Since the 1940s, cognitive scientists have followed the methodology to reproduce and investigate the mechanisms of cognitive functions by means of a kind of artificial simulation known as *model-oriented* simulation (Cordeschi 2002). In recent years, a number of research programs whose goal is to reproduce the entire human brain seem to have changed the methodology, by introducing the *data-oriented simulation* (Datteri, Laudisa 2016). I address some of the issues related to the evolution and epistemological implications of simulative methods in cognitive science. Furthermore, I will analyze the difference between these methodologies, in order to investigate the possibility of including the *data-oriented simulation* in the *model-oriented simulation*.

Francesco De Bei
PhD in Dynamical and Clinical Psychology, Dept. Dynamic and Clinical Psychology, Sapienza University of Rome, Italy.
Email: francesco.debei@uniroma1.it

Only Connect. The Contribution of Michael Tomasello to the Machine-organism Symbiosis

In recent years, the idea of a human-machine symbiosis able to lift individuals from burdensome tasks has undergone profound change. Not only because we can now rely on powerful personal computers, online databases and the Internet; but mostly because these technological tools have stopped being simple objects capable of performing mechanical tasks, and have become integral part of the environment. At the same time, also the concept of the body has undergone a radical revision. Both clinical and developmental research and cognitive science have led to rethinking the boundary between inside and outside, and to conceive of the mind as situated and simultaneously composed of internal and external processes. Moving from the evolutionary hypotheses developed in recent years by Michael Tomasello, this contribution focuses as well on how the meaning of a symbiosis between body and machines has been changing, as on the impact that an environment increasingly saturated of technology can have on mental functioning.

Mattia Della Rocca
PhD in History of Science, Dept. of Civilisations and Forms of Knowledge, University of Pisa; Temporary Fellow at Campus Bio-Medico University of Rome, Italy.
Email: mattia.dellarocca@gmail.com

From Brains as Machines to Machines as Brains. A Short Historical and Epistemological Reflection on the Simulation and "Reverse Engineering" of the Central Nervous System

21st century "big brain science" initiatives – especially the EU Flagship *Human Brain Project* – overtly aim to achieve a simulation of the central nervous system by "reverse engineering". However, a reflection is needed on the theoretical and epistemological assumptions which support this endeavour, and on the implications of adopting an engineering perspective in neuroscience for the contemporary Western discourse about the brain/mind system. In this contribution, I highlight the historical and epistemological origins of this peculiar perspective, focusing on some earlier electronic and engineering models of both the structure and the functioning of the brain – a paradigm shift that was tightly linked in neuroscience to the first attempts to develop a "neuro-technological transfer". Challenging the idea that the actual approach really represents a novelty in the panorama of neuroscience helps understanding the major criticisms to present-day brain simulations, and elucidating the role of extra-scientific factors in the development of this technological and scientific trend.

Simone Guidi
PhD in Philosophy. Lecturer at NABA (New Fine Arts Academy), Milan, and Membro Colaborador of Instituto de Estudos Filósoficos, Universidade de Coimbra, Portugal.
Email: simoneguidi@live.it

Mechanism "Prehistory" and the Strange Case of Cureau de La Chambre

This article deals with the concept of "mechanism" from a historical point of view, focusing on its relationship with the evolution of hylomorphism in the 17th century. I try to address the following questions: is mechanism structurally bound to materialism or does it rather represent a form of complete determinism, reconcilable with an "updated" version of hylomorphism? In the first part of the essay, I make the point that the very notion of "mechanism" must be clarified by means of a distinction between Boylean experimental mechanism and what Daniel Garber has called the "pre-history of the Mechanical Philosophy". My aim is to highlight how the deterministic (and nominalistic) hylomorphism developed in the 17th Century came quite close to mechanism. In this framework, I present the 'strange case' of Marin Cureau de La Chambre (1594-1669), which represents a characteristic compromise – based on the possibility of a not bodily extension – between a deterministic mechanization of the lower functions of the vegetative and sensitive soul and Campanella's panpsychism.

Fiorenza Lupi
PhD student in Philosophy, Dept. Philosophy, Sapienza University of Rome, Italy.
Email: fiorenza.lupi@gmail.com

The Organism, Maker of Machines

The contribution deals with Georges Canguilhem's philosophy of technique, addressing in particular two of its crucial aspects: *technique-science* (the technological problem) and *organism-machine* (the biological problem). Since the 1930s, Canguilhem develops a critique to the positivist theory that he synthetized as "knowing in order to predict and predicting in order to act". His argument is that the technical act is not a mere application of scientific principles, but rather the necessary condition for them: the theory is in fact the result of the reflection on the failures of the technical act. I also address the topic of the assimilation of the organism to the machine and of the use of this analogy in biology. Canguilhem in fact reverses the traditional relation between machine and organism: while it is true that the living being produces mechanical processes, it is however far more relevant that the machines, on their turn, are capable of adaptation, learning and interaction with the environment through processes of trial and error, based on features (organization and plasticity) once considered distinctive of the living beings. The paper aims at showing the philosophical consequences of this reversal.

Alessandra Passariello
PhD student in Philosophy, Dept. Philosophy, Sapienza University of Rome, Italy.
Email: alessandra.passariello@uniroma1.it

Early Organicism and its Juggling Machines: Further from Nature, Closer to Organisms

The article analyses the origin of the machine-organism debate with respect to the rise of the Organicist movement in the late 1920s. I argue that a better understanding of the Organicist divide between organisms and machines can be only grasped if we extend our analysis to the broader domain of relations between organisms, machines and natural phenomena. In particular, by analyzing two of von Bertalanffy's earlier texts (1933, 1952), I attempt to show that early Organicism assumed an epistemological proximity between machines and natural systems, while the machine-organism divide was rather instrumental in rejecting the physical-chemical reducibility of organisms. I highlight the role of cybernetics as the epistemological turning point in the loosening of the relationship between machines and natural

phenomena, and how this loosening paved the way to the rapprochement between organisms and machines.

Francesco Restuccia
PhD student in Philosophy, Dept. Philosophy, Sapienza University of Rome, Italy.
Email: francesco.restuccia@uniroma1.it

Mixed Systems and Interplay. Norbert Wiener meets Walter Benjamin

The article explores Wiener and Benjamin's views about the interaction between human beings and machines, highlighting some similarities. Despite their different areas of interest, they both recognize two kinds of technology, according to the sort of interaction each entertains with human beings: an older one based on mastery and a second one based on interplay. The latter may eventually lead to «systems of a mixed nature, involving both human and mechanical parts», a new collective «which has its organs in the new technology».
The emphasis the two authors share on interaction supports a view of automation that does not delegate human concerns to machines, but encourages the use of the new extended field of action to face these concerns in a new playful fashion together with the apparatus. In a society where machines are more and more able to pursue their own aims, human beings must therefore become even more responsible in setting and preserving the specifically human aims.

Luca Tonetti
Ph.D. in Philosophy, Dept. Philosophy, Sapienza University of Rome, Italy.
Email: tonetti.luca@gmail.com

Machines and Diseases: Giorgio Baglivi and his Mechanistic Physiopathology

The Croatian physician Giorgio Baglivi (1668-1707), professor of anatomy and surgery and, from 1702, of theoretical medicine at the *Studium Urbis*, is traditionally considered a leading representative of the Italian “iatromechanism”. Iatromechanism (or iatrophysics) is a 17th-century medical tradition, which attempts to explain normal and diseased states of the body in mechanistic terms, according to the laws of physics. In a famous passage from *De praxi medica*, quoted by Canguilhem in *Machine et or-*

ganisme, Baglivi compares the parts of the body to different kinds of physical devices, such as retorts, hydraulic pipes, springs, ropes, and so on. However, despite this radical reductionism, Baglivi also realizes that diseases cannot be entirely reduced to physics, the morbid states of the body being something completely different from malfunctioning of machines. Baglivi in fact rejects the use of mechanical devices for the treatment of diseases; in his view, the improvement of medicine only depends on the practice itself.
Therefore, Baglivi recognizes, albeit not explicitly, the limits of iatromechanism in pathology, proving that this category needs a thorough revision. In this paper, I examine the role of mechanics in Baglivi's medicine, by analysing, in particular, the problem of the relationship between solids and fluids in the body.

Valentina Trombetta
PhD student in Cognitive Science, Graduate School of Robotics, University of Freiburg, Germany.
Email: rinaldo89@alice.it

The Role of Emotions and Intrinsic Motivations on Decision-Making: a Comparison between Natural and Artificial Systems

Decision-making is a cognitive process that involves a choice, that is the capacity of a system to select one of alternative options. This is important for the survival of animals in their environment, but also for human beings facing moral dilemmas and finally for robots, to survive in a specific environment as well as to simulate real human behaviours. Different levels of analysis are necessary for developing a complete explanation of this process: the neurological level of the brain areas involved, the psychological one for the role of emotions and intrinsic motivations on choice, the philosophical analysis of the intentions as cause of a choice and/or an action, and robotics to compare the behaviour of natural and artificial systems. All these features are linked together within a theoretical enactive and dynamical framework. I will try to show how current research addresses the scope of comparing the role of emotions and intrinsic motivations in natural and artificial systems.

Index of Names

Filosofia e saperi - 9
Sconfinamenti tra i saperi umanistici e le scienze della vita
Crossing borders between the humanities and the life sciences

Wired Bodies

New Perspectives on the Machine-Organism Analogy

The scope of this series is to foster research devoted to a dynamical representation of the relationship between human sciences and life sciences and practices, and to stimulate new theoretical perspectives capable of supporting the communication and interaction between different disciplinary fields and thought styles.

The machine-organism analogy has played a pivotal role in the history of Western philosophy and science. Notwithstanding its apparent simplicity, it hides complex epistemological issues about the status of both organism and machine and the nature of their interaction. What is the real object of this analogy: organisms as a whole, their parts or, rather, bodily functions? How can the machine serve as a model for interpreting biological phenomena, cognitive processes, or more broadly the social and cultural transformations of the relations between individuals, and between individuals and the environments in which they live?
Wired bodies. New perspectives on the machine-organism analogy provides the reader with some of the latest perspectives on this vast debate, addressing three major topics: 1) the development of a 'mechanistic' framework in medicine and biology; 2) the methodological issues underlying the use of 'simulation' in cognitive science; 3) the interaction between humans and machines according to 20th century epistemology.

Contributors:
Francesco Bianchini, Elisa Binda, Giuseppe Boccignone, Nicole Dalia Cilia, Francesco De Bei, Mattia Della Rocca, Simone Guidi, Fiorenza Lupi, Alessandra Passariello, Francesco Restuccia, Luca Tonetti, Valentina Trombetta.

Nicole Dalia Cilia is PhD in Philosophy and History of Philosophy at Sapienza, University of Rome. Her research interests concern the epistemology of cognitive models, computational neuroscience and experimental psychology.

Luca Tonetti is PhD in Philosophy and History of Philosophy at Sapienza, University of Rome. His main research interests concern the history of science and medicine in the 17th and 18th centuries, with a particular focus on the history of medicine in Italy.

Cover illustration:
Vito Fabrizio Brugnola, *Singularity* (2016)

€ 12,00

ISBN 978 88 8080 236 5