Contemporary Systems Thinking

Gianfranco Minati *Editor*

# Multiplicity and Interdisciplinarity

## Essays in Honor of Eliano Pessa

Springer

# Contemporary Systems Thinking

*Series Editor:*

Robert L. Flood
Maastricht School of Management
Maastricht, The Netherlands

Contemporary Systems Thinking is a series of texts, each of which deals comparatively and/or critically with different aspects of holistic thinking at the frontiers of the discipline. Traditionally, writings by systems thinkers have been concerned with single theme propositions like General Systems Theory, Cybernetics, Operations Research, System Dynamics, Soft Systems Methodology and many others. Recently there have been attempts to fulfill a different yet equally important role by comparative analyses of viewpoints and approaches, each addressing disparate areas of study such as: modeling and simulation, measurement, management, 'problem solving' methods, international relations, social theory and last, but not exhaustively or least, philosophy. Bringing together many sources yields several achievements, among which is showing a great diversity of approaches, ideas and application areas that systems thinking contributes to (although, often with difficulties unresolved). There is a need for a series of books, each focusing in detail on the study areas mentioned above. While modeling and simulation are served well in the scientific literature, this is not the case for systems thinking in management, 'problem solving' methods, social theory, or philosophy to name a handful. Each book in this series makes a contribution by concentrating on one of these topics.

More information about this series at http://www.springer.com/series/5807

Gianfranco Minati
Editor

# Multiplicity and Interdisciplinarity

## Essays in Honor of Eliano Pessa

Springer

*Editor*
Gianfranco Minati
Italian Systems Society
Milano, Milano, Italy

ISSN 1568-2846
Contemporary Systems Thinking
ISBN 978-3-030-71876-3    ISBN 978-3-030-71877-0 (eBook)
https://doi.org/10.1007/978-3-030-71877-0

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In memory of Professor Eliano Pessa
September 19, 1946–March 22, 2020

*Eliano Pessa, Theoretical Physicist, was a Full Professor of General Psychology and Cognitive Modeling at the University of Pavia, Italy. He has already been Dean of the Department of Psychology and the Inter-departmental Research Center on Cognitive Science in the same university. He was also previously Associate Professor of Artificial Intelligence at the University of Rome "La Sapienza" Faculty of Psychology. He was the author or coauthor of ten books and many papers in scientific journals, books, and proceedings of international* *conferences. His scientific research interests included quantum theories describing the human brain's operation, computational neuroscience, artificial neural networks, modeling emergence processes, quantum field theory, phase transitions in condensed matter, human memory, visual perception, decision-making, and statistical reasoning. As an expert mountaineer, he has climbed mountains on various continents.*

# Preface

This book has been somewhat unconventionally organized. Usually, edited books are organized around a dominant, characterizing topic elaborated by the contributions from authors. This structure is the case for the proceedings of conferences. In the case of this book, the dominant subject is the human, cultural, and scientific contribution of a particularly extraordinary person, Professor Eliano Pessa, who recently passed away. It is a kind of book of proceedings of a virtual conference in his honor, the AIRS[1] conference of which he was the fundamental contributor. A limited part of the music that the orchestra of his life played is artificial intelligence, bio-systemics, cognitive science and psychology, quantum physics, systems science, and alpinism. These were never separate disciplinary issues, but rather, some of the coherent dimensions of interest in his life.

It is a matter of ongoing mutual, reciprocal interpretations and representations of approaches, concepts, problems, and solutions, considering shared contextual meanings. It is a matter of multiple contextual meanings *whose coherence is given by the fact of being lived, in this case, being lived by Eliano*. True interdisciplinarity and usage of nonequivalences can be described, in our case, among Eliano's life's interests. He lived and not only theorized interdisciplinarity and multiple dimensions. We remember how he considered it unacceptable *doing one only thing in life*, and we consider proudly the one that we have experienced with him. Indeed, Eliano practiced different ways and inspired most of us to do the same.

The contributors to this work had a considerable challenge, to be inspired by, rebuild, and retrace such networked interests lived by Eliano from a personal, cultural, and scientific perspective. The contributors were challenged to consider, discuss, interpret, and represent the multiplicity and interdisciplinarity experienced, lived, and applied by Eliano. Most of the authors lived with Eliano, the scientific story of the Italian Systems Society[1] of which he was the foremost scientific supporter, contributor, and mentor.

---

[1] Associazione Italiana per la Ricerca sui Sistemi, in English Italian Systems Society, see Postscriptum.

This book tries to honor such richness of Eliano's contributions. We abstain from any celebration that he would certainly not have liked. His life, scientific production, and mountaineer practice are common threads that authors elaborate upon, interpret, understand, and rebuild memories. We are called upon to create novelty, take risks, and not only to end up in contributions, the solidity of which has already been confirmed in the literature, and is perhaps even a little out of date.

We are happy and honored to have walked with him.

Now, a posteriori, we must have deserved him.

In honor and in memory of Eliano.

Milan, Italy
January 2021

Gianfranco Minati

# Post-scriptum

I founded the Italian Systems Society (AIRS) http://www.airs.it in 1996. The AIRS is a network of academicians, scientists, researchers, and professionals involved with systems research. The list of disciplines involved includes Architecture, Biology, Economics, Education, Engineering, Mathematics, Neurosciences, Medicine, Music, Philosophy, Psychology, and Physics.

The AIRS conferences have had distinguished open lecturers, including professors Arecchi, Haken, Kauffman, Klir, and Longo. The list of volumes of proceedings published includes:

Minati, G. (Ed.). (1998). *Proceedings of the First Italian Conference on Systemics, Apogeo scientifica*, Milan, Italy.

Minati, G., Abram, M., & Pessa, E. (Eds.). (2009). *Processes of emergence of systems and systemic properties. Towards a general theory of emergence*. World Scientific.

Minati, G., Abram, M., & Pessa, E. (Eds.). (2012). *Methods, models, simulations and approaches—Towards a general theory of change*. World Scientific.

Minati, G., Abram, M., & Pessa, E. (Eds.). (2016). *Towards a post-Bertalanffy systemics.* Springer.

Minati, G., Abram, M, & Pessa, G. (Eds.). (2019). *Systemics of incompleteness and quasi systems*. Springer.

Minati, G., & Pessa, E. (Eds.). (2002). *Emergence in complex cognitive, social and biological systems*. Kluwer.

Minati, G., Pessa, E., & Abram, M. (Eds.). (2006). *Systemics of emergence: Research and applications*. Springer.

# Contents

# Eliano Pessa Tribute to Three Voices

**Giordano Bruno, Franco Eugeni, and Alberto Trotta**

**Abstract** A portrait of Eliano Pessa, in the form of thee voices tribute. Graduated in Physics, with post-graduate training at the mythical group of Professor Caianello, he became first Associate of Institutions of Mathematics, and then thanks to his systemic studies in psychology he obtained the chair of General Psychology. We briefly recall here its vast culture, its human characteristics, its propensity for dialogue, its scientific production in the most varied fields of knowledge. Finally, we like to emphasize his great love for the mountains, as soon as free from its multiple commitments has climbed the highest ones in various places in the world, demonstrating great preparation, resistance and passion.

**Keywords** Alpinism · Mathematics · Physics · Portrait · Psychology · Scientific interests · Teaching

## 1 Introduction

It is always difficult to draw a portrait of a person who is no longer there. Even more so when it comes to a friend. So, we thought of uniting our feelings towards him, in order to create a kind of three-voices singing, which would allow to describe, in some ways, not only the scientist Eliano Pessa, but also the man in its facets.

---

G. Bruno (✉)
Universitas Mercatorum, Rome, Italy
e-mail: giordano.bruno@unimercatorum.it

F. Eugeni
A.F.S.U., Giulianova, Italy

A. Trotta
I.I.S.S. "S. Caterina Amendola", Salerno, Italy

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_1

## 2 Contribution of Giordano Bruno

Warm, friendly and willing to listen, despite his countless commitments both as a scientist and as a man. A man who was eager to reach new heights, not only those of his beloved mountains but also in the fields of physics, mathematics and psychology. This is how I like to remember Eliano.

When I first met him, through my good friends Bruno Rizzi and Franco Eugeni, I was fascinated by his wide knowledge and his natural kindness.

It is always a very pleasure to think back to the meetings we held in Bruno Rizzi's office, together with Franco. We discussed *number theory*, *relativity* and *quantum theory*, *non-Euclidean geometry* and several other scientific topics.

*Teaching* was also one of the interests we shared and valued significantly, and Mathesis, an ancient and reputable Society for mathematical and physical sciences, provided a space for it to thrive.

Indefatigable and at the same time cheerful, insatiably curious and extremely energetic, Eliano was always willing to lend a hand when it came to organizing conferences or seminars.

And then suddenly, out of the blue, whenever he could, he would vanish. "I wonder where he ended up this time", I thought and, invariably, a postcard would come after a while, from the wonderful, distant places he travelled to.

Eliano later started a fruitful collaboration with Gianfranco Minati, founder and president of the Italian Association for System Research (AIRS), contributing significantly to the organization and sharing his knowledge, the key to the scientific value of his research.

Meeting Eliano was a turning point in my life, both cultural and social, as well as an enriching personal experience.

He had introduced Alberto Trotta, who had meanwhile become my friend, to AIRS. Alberto had introduced me to Systemics and invited me to attend one of the Conferences the Association holds every 3 years—the second, if my memory serves me well. The title was "Emergence in complex cognitive, social and biological systems" and it took place in Castel Ivano (TN) in 2001.

Almost 20 years have passed since my first AIRS Conference. Since then, I have always made an effort to attend and to take part in the other initiatives the Association promotes.

Each of them is an opportunity to learn and improve among friends, and it is Eliano who I have to thank for this.

Although we saw each other rarely, our friendship and intellectual exchanges were never affected, and for this, too, I owe him.

Thank you, my friend.

## 3 Contribution of Franco Eugeni

I met Eliano in the early ‘70s, while he was still studying physics at university. He was with one of my students—back then, I used to teach Complementary Mathematics in L’Aquila—who had come to talk about his thesis. I was immediately impressed by his grasp on mathematics, a command so remarkable that we ended up preparing his friend’s project together.

I met him again in the ‘80s, in Rome, teaching Mathematics at the Psychology Faculty. At the time, he was collaborating both with Professor Caianello’s team and with my good friend Bruno Rizzi. This latter partnership produced about 20 remarkable works on applied mathematics.

In 1986, Bruno Rizzi and I had both obtained a full professorship. Together with Eliano and Luigia Berardi, we sought to broaden our intellectual and professional horizons by exchanging ideas with mathematical economists.

I remember that one day Eliano visited me in L’Aquila and I gave him one of my German shepherd’s puppies.

Shortly thereafter, he moved to Pavia, where he held the chair of Psychology, a field where he excelled as much as he did in that of Mathematics.

It is quite hard to find something to say about Eliano. I remember his cheerfulness, which could sometimes puzzle those who didn’t know him well, and his devotion to his many interests, which allowed him to stand out while still remaining humble and spontaneous.

I only regret not having seen him in these last 10 years.

## 4 Contribution of Alberto Trotta

Eliano Pessa was born in Portogruaro (Venice) on 19th September 1946, but grew up in Rieti, where his father, Professor Giuseppe Pessa, had accepted a job.

Eliano graduated in Physics at the University of L’Aquila and specialized in Astrophysics and Theoretic Physics at the University of La Sapienza, in Rome, with a thesis about “The theory of a quantized scalar field in a Bianchi I-type universe”. He was associated with the FUCI (Federazione Universitaria Cattolica Italiana), chaired by Don Lorenzo Chiarelli. It was during this time that he met his future wife and started a family.

He then began teaching, first at the ITIS (State Industrial Technical Institute) in Rieti and then at the University of La Sapienza.

Physics, however, was not the only field he excelled in: while working in Rieti and Rome, he kept in touch with a local group of amateur alpinists like himself and embarked on several climbing enterprises, reaching peaks such as Vulcan Pissis (Argentine Andes, 6862 m), Vulcan Parinacota (Chilean Andes, 6342 m), Pik Lenin (Kirgizistan Pamir, 7134) and Mount Muztaghata (Chinese Pamir, 7546).

Eliano was a remarkable person, and a scientist most competent in various fields. He took on several academic positions. Between 1977 and 1987, he taught Mathematics at the University of La Sapienza, Rome, first at the “Magistero” Faculty and then at the Psychology Faculty; after that and until 2000, he taught Theory of Systems and Artificial Intelligence. Between 1992 and 2000 he was a member of the Scientific Committee of the International Institute of High Scientific Studies E.R. Caianiello, Vietri Sul Mare, Salerno. In 2000, he became Professor of Psychology at the University of Pavia, Letters Faculty; that same year, he joined the board of directors of AIRS (Italian Association for Systemic Research). Between 2002 and 2003 he taught “Modelli di Reti Neutrali” (Neural network models) at the SAFI (Scuola Avanzata di Formazione Integrata). Between 2006 and 2009 he was the Director of Interdepartmental Centre of Cognitive Science, at the University of Pavia, where, in 2010, he started teaching Cognitive Processing Models.

He was also a referee for multiple scientific journals, such as the International Journal of General Systemics and the International Journal of Theoretical Physics.

He received many acknowledgements throughout his career, such as the “Majorana Field”, in 2008, awarding his work “Phase Transitions in Biological Matter” (Pessa, 2007).

His research dealt with a large variety of topics, combining his competence in physics and mathematics with his enthusiasm for cognitive sciences. With over 200 published works, Eliano explored the theory of quantum mechanics, neural networks, robotics, artificial intelligence, theoretical and experimental studies on long-term memory, analysis of global and local factors of visual perception, focusing and categorization process models, quantum field theory, quantum computing, quantum memory models, systemic and self-organization models in complex systems.

He also collaborated with many colleagues, such as the president of AIRS, Gianfranco Minati, former president of Mathesis Bruno Rizzi, Mario Abram and M.P. Penna.

Among the many notable works he published, we particularly mention contributions such as (Pessa, 1985a, b, 1992, 1993, 2005). Among co-authored works we mention (Minati & Pessa, 2018; Penna & Pessa, 1994; Pessa & Rizzi, 1987, 1988; Pessa & Trotta, 2008). Among the edited proceedings we mention (Minati & Pessa, 2002; Minati et al., 2006, 2008, 2015, 2018).

A dedicated scholar, invested in his work, Eliano also lived a full life and passed away on the 22nd March 2020, following a long illness.

His is greatly missed, as great is the feeling of emptiness his absence brings along. In an article for the Italian newspaper *Il Messaggero* (26 March 2020), Arnaldo Millesimi wrote: “Eliano always walked on the tip of his toes, in order not to tread on others, and on the tip of his toes he went away”. This is, to me at least, a most relevant trait of his personality.

Eliano was a dear friend, cordial, considerate, forward-looking, determined and remarkably competent. He was a mentor for entire generations of scholars, and a scientist ahead of his time, exploring theories that are still in the forefront to this day, such as the statistics of neural networks.

His legacy shall never be forgotten.

**Acknowledgments** We wish to thank Elisabetta Passavanti for her precious contribution to the English translation.

## References

Minati, G., Abram, M., & Pessa, E. (Eds.). (2015). *Toward a post-Bertalanffy systemics*. Springer.

Minati, G., Abram, M., & Pessa, E. (Eds.). (2018). *Systemics of incompleteness and quasi-systems*. Springer.

Minati, G., & Pessa, E. (Eds.). (2002). *Emergence in complex, cognitive, social, and biological systems*. Kluwer.

Minati, G., & Pessa, E. (2018). *From collective beings to quasi-systems*. Springer.

Minati, G., Pessa, E., & Abram, M. (Eds.). (2006). *Systemics of emergence research and development*. Springer.

Minati, G., Pessa, E., & Abram, M. (Eds.). (2008). *Processes of emergence of systems and systemic properties—Toward a general theory of emergence*. World Scientific.

Penna, M. P., & Pessa, E. (1994). *La rappresentazione della conoscenza, introduzione alla psicologia dei processi cognitivi*. Armando.

Pessa, E. (1985a). *Algoritmi, automi, reti nervose*. Kappa.

Pessa, E. (1985b). *Stabilità e auto-organizzazione, introduzione alla teoria dei sistemi*. Veschi.

Pessa, E. (1992). *Intelligenza artificiale teorie e sistemi*. Bollati Boringhieri.

Pessa, E. (1993). *Reti neurali e processi cognitivi*. Di Renzo.

Pessa, E. (2005). La Matematica per la vita Artificiale. In G. Bruno, & A. Trotta (a.c.). *Conoscere Attraverso La matematica: Linguaggio, Applicazioni e Connessioni Interdisciplinari* (pp. 455–467). Mathesis.

Pessa, E. (2007). Phase transitions in biological matter. *Electronic Journal of Theoretical Physics, 4*(16), 167–230.

Pessa, E., & Rizzi, B. (1987). Sulla formulazione matematica dei modelli economici di tipo morfogenetico. *Rivista di Matematica per le Scienze Economiche e Sociali, 10*(1/2), 65.

Pessa, E., & Rizzi, B. (1988). Noise-induced transitions in economic dynamical models. In M. Galeotti, L. Geronazzo, & F. Gori (Eds.), *Non-linear dynamics in economics and social sciences* (pp. 259–269). Pitagora.

Pessa, E., & Trotta, A. (2008). Matematica e sistemica. *Periodico di matematiche, 3*, 105–123.

# The Role of Information and Communication Technologies in Human Interactions

**Mario R. Abram**

**Abstract** The development and diffusion of Information and Communication Technologies (ICT) increased the availability of a large number of communication channels. Some big problems arise when these new tools operate pervasively in our lives. The evolution of these technologies and their development, with the help of uncountable applications, are influencing and modifying our behavior. Unfortunately the respect of human rights is based again on the application of regulations and laws that are formally structured around the human beings. The application of laws is localized into the domain of the national law systems. Furthermore the distance between the quick evolution of the new technical applications and the slow update of the law systems increases constantly. The artificial tools that compose ICT applications are now again out of any regulatory system. In particular the experiences imposed by the diffusion of COVID-19 virus show the great help given by the application of ICT to contrast and react the pandemic diffusion. At the same time the risks to loss the respect of human rights emerge with great evidence.

**Keywords** Communication channel · Covid-19 · Human interactions · Human rights · Information and communication technologies · Interacting systems · Pandemic · Regulations and laws

## 1 Introduction

Investigating the impact of new Information and Communication Technologies (ICT) on human lives, we were attracted by the impact that the developments of new technologies may have over the spectrum of the human lives.

In a previous work (Abram & Pessa, 2019) we examined some of the possible consequences that arise developing and using the real applications of Information

---

M. R. Abram (✉)
AIRS (Associazione Italiana per la Ricerca sui Sistemi), Milan, Italy

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_2

and Communication Technologies (ICT). In particular it becomes important to evaluate the impact of ICT on the application of the human rights. The connection with regulations and laws appeared us as a crucial point.

We need to develop a lot of investigation work in order to set the problems, to define priorities and to propose solutions. Additional researches are necessary in order to deeply investigate all the implications of this approach.

Now this appears more evident because the pandemic COVID-19 is showing its devastating impact on all our activities. Probably it is necessary to develop new strategies, new models, for gaining the knowledge necessary to find an escape path from this very serious situation. Many severe and rigid health protocols were activated in order to protect people from the pandemic diffusion.

In the mean time the ICT technologies showed their great impact as a powerful alternative path useful to overcome the loss of communications and physical interactions between individuals and organizations.

In this paper we recall some considerations about ICT. In particular following the steps previously investigated (Abram & Pessa, 2019), we will consider and discus some interesting points.

The interactions between the development of Information and Communication Technologies (ICT) and the problem to control their applications is recalled (Sect. 2). Then the developments of new technologies ask for the settlement of new regulations and laws. Some implications are shortly investigated (Sect. 3).

Considering the importance of the natural and artificial contexts, we show how the role of communications between humans may be exercised by means of natural and artificial channels (Sect. 4).

The aspects previously investigated appear very actual. During this year of COVID-19 pandemic, it emerged the real need to solve the new and critical regulations aspects (Sect. 5).

Some remarks and open problems show the need for further investigations (Sect. 6). Finally some conclusive considerations close the paper (Sect. 7).

## 2 Information and Communication Technologies

In the actual pandemic context, the need to quickly gain useful results asks to accelerate the development of new medical and biological treatments.

The emergency conditions determined a rapid deployment of confinement strategies in order to reduce the physical interaction between people.

The contemporary need to reduce physical interactions between individuals accelerated the development and the use of more pervasive and powerful instruments based on ICT. In particular ICT services constitute the kernel of new recovery tools with the goal to mitigate or substitute the strong reduction of physical interactions between people.

The urgency to save the life of human people and the need to move quickly ask for reducing the response time of people, organizations, countries and governments.

The present situations recall the considerations we reported into our last contribution (Abram & Pessa, 2019); many consequences are now valid again and often are amplified.

## 3 The Role of Regulations and Laws

With Prof. Eliano Pessa we wrote the paper "*Information, Communications Technologies and Regulations*" (Abram & Pessa, 2019) with the goal to gain a better position useful to investigate the great changes we are facing with.

It was a tentative to find a reference point from which the different approaches to the problems may be evaluated. With the help of a systemic approach it is possible to gain a point of view from which larger perspectives for evaluating the real application of human rights become available.

Doing so we face with philosophy, history, power, economy, cultures, traditions, and all the other numerous aspects that may affect and characterize the evolution of mankind.

This is the ideal situation in which the constitutional guaranties and the deployment of human rights are sacrificed to collective and personal health needs. It is a very serious situation in which the urgency to operate conflicts with the respect of human rights.

The economic consequences of this situation may constitute further elements that supply dangerous feedbacks that may degrade the mutual interactions between the different elements of the societies. Then the basic human rights often are dangerously at risk.

Following this approach the respect of human rights gained a particular attention. In particular same documents became naturally the starting points to begin the analysis and the building of regulations and laws. We consider very important the following documents:

- United Nations, "*Universal Chart of Human Rights*" (United Nations, 1948);
- European Union, "*Charter of Fundamental Rights of the European Union*" (European Union, 2016b).
- In particular for data protection the European Union published the document "*General Data Protection Regulation*" (European Union, 2016a).

These documents constitute a reference guide for the building of future laws and regulations. They constitute a reference by which it is possible to test the coherence of declared human values and the effective application of human rights. With the help of these documents we may start to test the coherence and the value of their application to human relations.

## 4 The Communication Environments

The communications between humans may develop into two contexts:

1. A *natural context* involving the direct relations between human beings, and
2. An *artificial context* in which the human communications are mediated with the help of artificial tools.

### 4.1 The Natural Context

With natural context we define an environment in which the human elements interact each other into the natural environment (in biological, chemical and physical conditions). In that environment they may develop their investigations and increase their knowledge.

In this context, with a very simple schema, two human elements $H_1$ and $H_2$ interact by means of action relations $A_{12}$ and $A_{21}$ (Fig. 1a). When an action $A_{12}$ is incomplete the interaction is modified (Fig. 1b). When the actions $A_{12}$ and $A_{21}$ disappear the human elements $H_1$ and $H_2$ are isolated.

It is the context in which the human relations develop and find fulfillment. We all interact with our fellows by speech, touch, smell and seeing according to numberless modalities. On this basis the highest levels of interaction develop the possibility to use languages, psychology, abstraction, theories, etc.

In this situations the social relations manifest themselves and develop on the physical level contributing in great quantity to build the entire spectrum of interpersonal relations. At this level the communication channels operate the information transfer by means of chemical, biological and physical interactions.

### 4.2 The Artificial Context

During the evolution of humankind the interactions between people took place with the help of many media, opening so new types of channels that give the

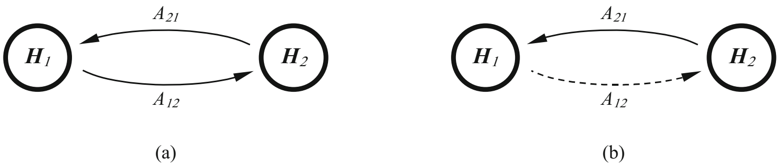


**Fig. 1** Simplified structure of interactions between two humans elements ($H_1$ and $H_2$): (**a**) two connected human elements; (**b**) two partially connected human elements

possibility to enlarge the range of human communications and activities. New tools based on communication and information technologies enlarged greatly the range of interactions between people.

We can consider an artificial context that is an extension of natural context with the addition of interactions and tools that multiply the availability of communication channels between the human elements. In this context the action domain of human elements is drastically enlarged. Now the individuals may operate on the extended range of a potentially worldwide space of operation.

The communications between $H_1$ and $H_2$ now can take place with the help of artificial tools $T_1$ and $T_2$. Then the communications between $H_1$ and $H_2$ are possible on the availability of redundant channels (Fig. 2a). Two paths are possible: (1) a direct paths between $H_1$ and $H_2$ and (2) a composed path involving $H_1$ and $H_2$ with $T_1$ and $T_2$.

The artificial context becomes the solution to maintain communications between individuals $H_1$ and $H_2$ by means of a series of new channels implemented with the help of ICT technologies. These channels can bypass the interruption of the traditional communication channels between individuals.

When the direct relations between humans $H_1$ and $H_2$ disappear (Fig. 2b), the communications between $H_1$ and $H_2$ are now possible. Using the tools $T_1$ and $T_2$ an alternative and different path becomes possible.

In particular the communication between $H_1$ and $H_2$ is possible by means of the following chains and relations (Fig. 2a):

1. A *direct path*: along the chains ($H_1H_2$) and ($H_2H_1$) respectively with the relations ($A_{12}$) and ($A_{21}$).
2. A *composed path*: along the chains ($H_1T_1T_2H_2$) and ($H_2T_2T_1H_1$), respectively with the relations ($A_{11}C_{12}C_{22}$) and ($A_{22}C_{21}C_{11}$).

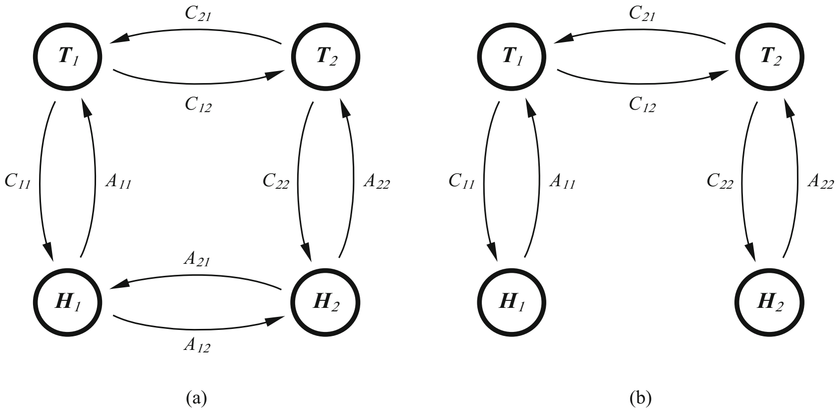


**Fig. 2** Simplified structure of the main interactions between two human elements ($H_1$ and $H_2$) and two artificial tools ($T_1$ and $T_2$): (**a**) two human elements are connected directly and by means of two tools; (**b**) two isolated human elements are connected by means of two tools

## 4.3 Interactions Between Natural and Artificial Contexts

The artificial tools are used to increase the number of channels for exchanging information between human elements and systems.

Now we can consider two separated spaces: (1) the space SH of human elements and (2) the space ST of the artificial tools.

The communication paths between $H_1$ and $H_2$ are possible by means of multiple channels (Fig. 3a). The space of tools ST may be composed by deep structures not directly seen by $T_1$ and $T_2$. Instead the space SH is composed only by the human elements $H_1$ and $H_2$ and their connections. We consider that the space SH is a natural space, while ST is an artificial space and all its connections are artificial.

When the elements of space SH are not connected directly, they interact each other only by means of the artificial elements of space ST (Fig. 3b). In particular when the chain of artificial tools is the unique communication channel between two human elements, the "owners" of communication tools can "control" the information flow between the human elements $H_1$ and $H_2$.

If one controls the elements (artificial tools) in ST, he can effectively manage the flow of information between the isolated elements of SH, especially when the alternative communication channels are closed.

As a consequence it is important to encourage the building and maintenance of communications between human elements and to perfect and optimize their use.

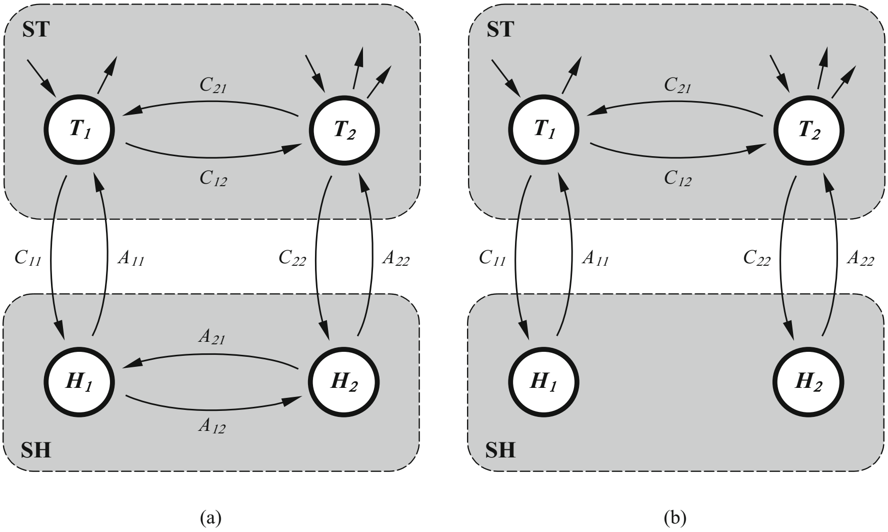


**Fig. 3** Structure of the main interactions between the spaces SH and ST composed respectively by two human elements ($H_1$ and $H_2$) and two artificial tools ($T_1$ and $T_2$): (**a**) two humans are connected directly and by means of two tools; (**b**) two isolated humans are connected by means of two tools

Redundant channels, as additional communication paths, are useful to check the correctness of transmitted and received information.

## 5 The Actual Pandemic Context

On the pandemic context we may develop some considerations.

- This is the situation people experienced in the past centuries. The physical contact was the most important diffusion factor for epidemic.
- The constraints imposed by the pandemic involve the entire world. Apparently no country appears free from these problems.
- The pandemic strongly reduced or interrupted the direct communication channels between the individuals. All the activities based on personal and direct contacts were reduced or interrupted.

In these situations it is important the availability of artificial communication channels by which many alternative paths may be activated. Then the interactions between human elements may continue with the help of artificial tools without exposing people to dangerous physical contacts.

Now many multilevel artificial tools that operate on a worldwide scale are available. It is possible to observe the emergence of the power of artificial tools that, out of law, multiply the availability of communication channels and enlarge the domain of the applications.

They furnish the solution to many communication problems, but effective regulations and laws do not rule them. Regulations and laws are now applicable only in a little set of countries, into their own domain of application that coincides with the domain of sovereignty of each nation.

A possible alternative is to regain the role of human elements that find their role of special relations between humans and gain the choice, ability and the courage to be in front of the new technologies without submission. It is necessary to find the courage to confront with new technologies and to identify their most appropriate deployment. The application domain of regulations and laws is possible only at human level.

Then the humans must gain again their role. Even in pandemic, during isolation, the tools may help to “meet” other people at different levels.

It is necessary to evaluate the need to regain a path for rebuilding human relations, enriching so role of human elements. It is as going back again from the conditions of isolated human element connected only by artificial tools (Figs. 2b and 3b) to a position in which the role of human elements returns to be central and the original multiple connections are restored. In these cases the communications between individuals return to be important, even if artificial tools are present into the transmission chains (Figs. 2a and 3a).

## 6 Remarks

The considerations developed into a previous paper (Abram & Pessa, 2019) are now applicable to today situations. The COVID-19 pandemic changed drastically the interactions between people. The dramatic consequences in terms of lost lives, economic crisis, social constraints and urgency planning are evident. All human activities were involved and stressed while the traditional strategies are showing their limits.

Some key points appear important in order to identify the goals, to define the priorities of the activities and to plan their realization.

- A chain of artificial tools may become the unique communication channel between people. The need to use these communication channels push to develop and standardize new communication protocols that progressively may drastically reduce the possibility to communicate directly between the persons. The risk to impose only an artificial communication may progressively reduce the direct natural communications.
- Our way of life was forced to change in order to confront with the need to survive this dramatic situation. The goal to reduce the risks of virus diffusion asks for a strong limitation of personal freedom and moving. It is necessary to acquire a strategy in order to reduce the risk of contagion and contemporary to develop and support the human activities. How much time a "frozen" society can survive?

  Information and Communication Technologies show how the new ICT tools and instruments help to reduce the physical interactions between people.
- When the natural interactions between human beings are interrupted, the artificial communication tools become the unique available channels interconnecting people. These aspects became more evident when, as a reaction to the diffusion of pandemic, the physical interactions between people were strongly reduced.

  In many activities the availability of redundant and artificial communication channels became the only possibility to interact each other. In the meanwhile the traditional human communications progressively reduced and often they were interrupted choosing to operate mainly with artificial channels.

  Then the pandemic amplified and accelerated a trend just present in our societies.
- A crucial point is that our law systems are built and evolved to regulate the human interactions. The artificial tools operate into a worldwide space; their ranges overcome the application domain of the human law systems. For this reason the more critical aspect is the fact that the artificial tools often are not subject to national norms or regulations because they operate outside the limited range of application of that norms or regulations.
- Usually the artificial channels are not a one to one connection system but may interact with a great variety of unidentified tolls interconnected with actors that remain unknown to the users. As shown in Fig. 3b it means that two people submitted to the constraints activated for reducing the effects of pandemic may

interact "only" by mean of artificial tools that often are not subject to any law or regulation because they can operate outside of the national borders.

If someone may control the artificial tools, he can control the information flowing through those tools. The connections with "power" become evident. Those that control the flow of information need to control the artificial communication channels.

- The pandemic is an amplifying factor that accelerates the phenomena just present in our societies. and may become dominant if we do not choice witch type of relation we want to build. Using Communication and Information Technologies (ICT) without adequate laws and regulations we may enter into a "risk space" in which the "game rules" are not chosen by the users and remain unknown.
- In building relations it is useful to maintain a multiple path approach. So it is convenient to maintain active some human paths, in order to develop a control strategy of our relations. It should be useful to avoid we delegate our communications only to artificial tools. It is necessary to avoid that human elements become only a "terminal" or an "artificial element" connected only to artificial tools.
- The experience of pandemic forced to emerge the recovery strategies inspired and settled by our conception of human elements and of artificial tools. Two possibilities may be considered: (1) the human elements maintain the control on artificial tools; (2) the artificial tools drive the behavior of human elements.
- The choice of the approach defines the strategies that we want to apply. The emergency strategies bound the role of human elements submitting them to the algorithms of the artificial tolls. This must be limited in time in order to return to more conscious choices. Humans should regain their degree of freedom.
- The availability of multiple communication paths develops the ability to evaluate different possibilities and to operate the best choices. For this reason it is necessary to develop the ability to collect information, to evaluate the available data, to develop a communication strategy, to build a path to the goal and to operate the best choices. In other words it is necessary to gain again the capability to "operate a choice".

## 7 Conclusion

In this period of pandemic many new problems appear and ask for a solution.

Between the available tools, the characteristics of Information and Communication Technologies (ICT) appear of crucial importance. With ICT we saw how many alternative paths are available to overcome the difficulties that are present in all the human activities.

Even in this situation it is important to evaluate the strategies that may conjugate the need of urgency and quick reaction with the best inspiring principles and operating actions.

In any case we think it is necessary to gain a better position from which, as men, we may evaluate the problems from the point of view of human rights.

**Acknowledgement** This paper is dedicated to the memory of Prof. Eliano Pessa.

## References

Abram, M. R., & Pessa, E. (2019). Information, communication technologies and regulations. In G. Minati, M. R. Abram, & E. Pessa (Eds.), *Systemics of incompleteness and quasi-systems* (pp. 221–233). Springer.

European Union. (2016a). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119, 4 May 2016 (pp. 1–88).

European Union. (2016b). Charter of fundamental rights of the European Union. *Official Journal of the European Union*, C202, 7 June 2016 (pp. 389–405).

United Nations. (1948). International bill of human rights. Universal declaration of human rights (General Assembly Resolution 217 A (III), Paris, 10 December 1948). *Official Records of the Third Session of the General Assembly, Part 1, 21 September – 12 December 1948. Resolutions* (pp. 71–77).

# Interdisciplinary Systems Thinking for a New Scientific Paradigm: Toward a Re-founding of Human Values

**Sergio Barile and Marialuisa Saviano**

**Abstract** In the face of an emerging scenario characterized by multiple overlapping trends that seriously question the future of mankind, this contribution envisions the need for a re-founding of human values in which process scholars are required to play a prominent role. Our focus is on the complexity of interdisciplinary thinking required for the development of a new widely shared paradigm. In particular, we believe that systems thinking can support interdisciplinarity as a common denominator of generalizable knowledge useful to build a shared thinking space that cross-cuts the boundaries of various disciplines. Within this context, the bases are delineated of a potential approach, favouring interdisciplinary convergence toward a shared vision of the future, via general systems thinking models developed within the Viable Systems Approach strand of studies.

**Keywords** Complexity · Generalizable knowledge · Human values · Information variety model · Interdisciplinarity · Scientific paradigm · Systems thinking · T-shaped model · Viable systems approach

## 1 Premise

Any analysis of the present time can only converge in noting various emerging 'identity' aspects of our era. First, a substantial alteration of existential rhythms: technology, in its broader sense, has to a great extent, accelerated all socio-economic processes. At the same time, compared to our predecessors of a mere few decades ago, we explicate activities that then would have required on average, tenfold or even 20 times more. Second, an extreme exaltation of specific knowledge: the

---

S. Barile (✉)
Sapienza University of Rome, Rome, Italy
e-mail: sergio.barile@uniroma1.it

M. Saviano
University of Salerno, Fisciano, Italy
e-mail: msaviano@unisa.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_3

development dynamics of knowledge have an exponential trend. A sentiment exists, regardless of the field of commitment, relative to the intellectual potential, and the cultural level, of being constantly in pursuit of new essential knowledge. Third, a prevailing nihilism: in many social contexts there is a growing sense of despair linked not to the contingent, but to the impossibility of imagining the near future. The collapse of every reference, from the religious to the political, to the most closely familiar, generates confusion and disorientation, producing the conditions for serious psychophysical disturbances (an element that has been exacerbated by the pandemic).

On the other hand, in some ways surprisingly, scientific and technological development advances producing innovation, even radical, with unusual speed. Robotics proceeds inexorably with a progressive replacement of human work, not only manual but now also intellectual, proving itself able to produce goods and services more efficiently at increasing quantitative and qualitative levels. Artificial Intelligence has long been forging indecipherable algorithms that continually regenerate, placing itself at the service of science and technology and of technique, with an unstoppable progression that suggests unpredictability and difficulty in controlling ways of using spaces and organizational and individual timelines. Augmented Intelligence, a phenomenon not yet well understood even by various streams of professionals, can be identified in the redefined concept of 'contemporary' individual, capable, by virtue of 'brain prostheses' (thus we could also define a simple search engine on the web), to qualify a sort of *mens extensa* generating a radical transformation of the human race (*Homo digitalis*). Genetic engineering, extremely close to freeing us from the imperfections that afflict *Homo sapiens*, seems to lead to a kind of biological engineering that opens up new evolutionary outlooks for possible hybrid life forms that will be able to populate our planet in the future.

It is difficult to comprehend the meaning of a potential temporal coincidence of these trends. The only unifying perspective is the one that allows the identification of a common underlying dimension: a sort of 'guiding principle' that leads to justifying the simultaneous presence of social trends apparently not attributable to any known interpretative scheme. This dimension must necessarily reconcile with that, due to the above highlighted 'identity' aspects, which we are witnessing in terms of a substantial dissolution of solid 'points of reference' that have always underpinned the behavior of mankind during the eras of historical development along the so-called 'path of civilization'. Priorities considered unavoidable, principles that appeared essential creak now under the pressure of legitimate expectations of greater well-being and quality of life. The same abovementioned aspects of development are beginning to question the 'strong beliefs' that typify the ethnic groups and human aggregations that have been present on the planet for centuries.

Such concepts all seem to converge towards the need for a re-founding of human values that takes into account the abovementioned trends in order to link them into a rediscovered purpose distinguishing the contemporary individual.

The modern world now appears strongly characterized by the well-known contrast and at the same time complementarity of two interacting tensions, the dipole between science and humanism: on the one hand, the incontrovertible power of science and its growing successes; on the other, the understanding of life as an experiential path, especially emotional, full of sentiment and value projections. A sort of ‘yin and yang’ in which the cults of acting and experiencing mix, intertwine, conditioning each other.

Thinking in terms of a new paradigm suitable for designing a possible evolutionary path in which to find explanations and predictions of current and future socio-economic scenarios appears urgent, no longer procrastinable. It is evident, however, that to achieve such a goal requires a titanic but necessary intellectual effort, in order to avoid the disintegration of founding principles that would lead to the loss of the ultimate ends of mankind.

Such a commitment should inspire scholars interested in what we could metaphorically define as the primary component of the ‘molecular’ fabric of civil society: ‘finalized’ organizations. It is a truism that just as an aggregation of atoms does not in itself qualify ‘bios’, given that cellular tissue is the basic element to which life can be traced back, in the same way the individual as such does not characterize sociality, because social being emerges when the aggregation of elements becomes a system, when the tribe becomes a community. Therefore, the desire for recovery of the values system cannot merely neglect the individual but rather, has to focus on their potential forms of socio-organizational composition.

Certainly, the various scholars involved in both scientific and humanistic disciplines would recognize that the distinctive feature of the human species, compared to other living species, consists in the fact that the former are a ‘collective’ species. It is precisely in their ability to work together the distinctive element that has enabled the survival of a species unsuited to the planet such as humans. As Harari maintains,[1] we are the only beings, compared to any other species, who need to integrate the natural environment to resist the cold and heat, and who need tools (artefacts) to build our ‘den’, to feed ourselves, to face enemies, etc. We would obviously be losers in a ‘single fight’ against any specimen of ferocious animal; conversely, by creating a community of a few dozen individuals, we become unbeatable with respect to aggressive fauna (such as lions and tigers), not to mention bacteria and viruses, by virtue of our ability to organize *collective* defence and planning against attack.

In the belief that those who study organizations aimed at the production of value in the broadest sense are most qualified to promote and coordinate debate directed at recovering a new explanatory paradigm capable of summarizing the trends reported above, we wonder:

*How should scholars think together about the foundation of a new paradigm? Are they equipped with the appropriate approaches to address such a challenging task?*

Promoting debate on the identification of constructs, schemes, categories, elements in short, which can be placed at the basis of a re-founding of valuesuseful

---

[1] Harari, Y.N., (2015). What explains the rise of humans? TED London, June 2015. https://www.ted.com/talks/yuval_noah_harari_what_explains_the_rise_of_humans?language=it

for recovering the ultimate meaning of life, is the aim of this essay. Accordingly, in the following sections, we first reflect upon the need of collective interdisciplinary thinking for such an innovative paradigm, focusing on the role of scholars and the way they can approach complex interdisciplinary thinking; then, we discuss the contribution of systems thinking as a common denominator for approaching interdisciplinary knowledge sharing; subsequently, we illustrate the use of the Information Variety and T-shaped knowledge models, developed within the Viable Systems Approach strand of managerial studies of systems thinking. Our aim is to approach and organize an interdisciplinary knowledge co-creation context, highlighting how the proposed approach can result in promoting a re-founding of values for the development of a shared paradigm.

## 2 Collective Interdisciplinary Thinking for a New Paradigm

As outlined in the introductory premise, nowadays human beings are facing trends whose directions and arrival points are completely unknown. At the moment, for example, we are living the context of the Covid-19 pandemic, a completely unexpected crisis that, despite huge governance effort, appears still out of control, dramatically impacting on all spheres of human life. Recovery plans are under discussion and about to be defined to face not only the specific health problem but also its serious social and economic consequences, envisioning a digitized and more sustainable world as a priority. No one, however, is sure about the real outcomes of such plans. No one is confident about the capability to make the right decisions. There is no clear-cut vision about the future we want. The capability itself to project our future is doubted. We do not really see where we want to go. Clearly, there is a widespread lack of finalities.

In a powerful globalized and technologically advanced world, an invisible entity has placed all human beings in the face of one of the toughest challenges ever. Intensive debate crowds the media about prospects and solutions to the global crisis in progress. In such a context, one apparently less relevant question hovers among disoriented people: *How will it be after Covid-19*? There is no clear feeling about whether and how mankind could eventually change after the pandemic. A pandemic, as the Greek origin of the word suggests, implies a condition affecting all (or nearly all) the people in the world. Never before has our generation had such an experience of something really ‘global’. Never before, in the globally digitized world have we experienced something that is really common to us all, that is, something that affects us *collectively*.[2] Paradoxically, neither globalization nor digitisation, so far, have

---

[2] The notion of collectivity has an interesting reference in the work of Minati and Pessa, entitled *Collective beings*, which focuses on the complexity of collective systems, such as social ones (Minati & Pessa, 2006).

placed us in front of a truly collective experience, showing all the contradictions that characterize the trends under focus.

*What has led us to such a paradoxical condition of uncertainty in a world ever more committed to solving human life problems, giving ever more 'certainties' to people?*

*What is the role of scholars and science in such a scenario? Will progressing technical, technological, scientific and practical knowledge still be enough or does science have to have more responsibilities?*

## 2.1 Scholars in Front of the Need of Complex Interdisciplinary Thinking

As repeatedly highlighted since the beginning of the last decade, the growing complexity experienced by decision makers requires the support of knowledge capable of reflecting experienced complexity (Badinelli et al., 2012; Barile et al., 2016a). The awareness of the need to rethink knowledge as a dynamic system, instead of a linear one, has led to promote *interdisciplinarity* as an approach expected to be more effective in the creation of knowledge useful to address complex choices (Klein, 1990; Brewer, 1999; Moran, 2010; Klein et al., 2001; Newell et al., 2001; Newell, 2007; Fadeeva et al., 2010; Frodeman et al., 2017; Bammer, 2017). The relationship between interdisciplinarity and complexity has become of increasing interest among scholars that recognize that the two concepts are deeply entwined (Klein et al., 2001; Newell, 2007).

Agreeing that mankind is facing trends difficult even to understand, science has the responsibility to provide not only answers but also guidelines, i.e. directions for thinking. All sciences, not only those 'disciplinarily' related, should feel involved in a reflection about the future of mankind and our current planet. Accordingly, it follows therefore, that one of the most urgent tasks for scholars is to promote and support a necessary re-founding of human values.

During the progressive development of mankind, science has always played (and claimed to play) a leading role. Indeed, science is at the core of the processes making possible equitable and sustainable development for all populations. Therefore, it is our belief that the time has come for a truly shared reflection about the future we want, that is, the ends we want to pursue. Such a view can shed new light on the meaning and role of interdisciplinarity.

Interdisciplinarity has long been practised producing important advancements of knowledge (Lélé & Norgaard, 2005). Nevertheless, debate about the need of inter- as well as transdisciplinary approaches is still in progress (Dube, 2021). In particular, as recently highlighted by Bammer et al. with reference to interdisciplinarity largely unrecognised is the need of expertise in (Bammer et al., 2020: 3), "(1) research integration in order to develop a more comprehensive understanding of the problem and potential ways to address it and (2) implementation of research to improve the situation." Although it is commonly recognized that expertise in inter- and transdisciplinary research is required to address complex societal and environmental problems, "there still remains little formal recognition of what that expertise is or the reward for contributing it to a research team's efforts" (Bammer et al., 2020: 3).

Moreover, the implementation of inter- and transdisciplinarity implies to address relevant requirements (Hadorn et al., 2008 : 19):

1. To grasp the complexity of the problems.
2. To take into account the diversity of scientific and societal views of the problems.
3. To link abstract and case specific knowledge.
4. To constitute knowledge with a focus on problem-solving for what is perceived to be the common good.

As Bammer et al. affirm, interdisciplinary (and transdisciplinary) approaches are necessary for developing "a more comprehensive understanding of the problem, identifying possible ways to address it, by integrating disciplinary and stakeholder perspectives, and for supporting implementation of that understanding into evident informed government policy, professional and community practice, business and social innovation and other measures." In this context, the Authors identify five particular challenges posed by complex problems (2020: 4):

1. Delimiting the problem.
2. Managing contested problem definitions.
3. Managing critical, unresolvable unknowns.
4. Managing real-world constraints on ameliorating the problem.
5. Appreciating and accommodating the partial and temporary nature of solutions.

While agreeing with Bammer et al. about both the need of inter- and transdisciplinary approaches for addressing complex issues and the recognition of a still limited attention on the ways knowledge (disciplinary and beyond) can be integrated and implemented, we return to and pinpoint a basic reflection with a slightly different view of the kind of knowledge required to address inter- and also transdisciplinarity: we believe that what is much needed is the capability of more general thinking with holistic views of the observed phenomena. In other words, a kind of *meta*-knowledge, i.e. knowledge a-specific and sufficiently general to unitarily direct the use of the more specific available approaches to deal with complexity (among the latters, for example, action research, complex project management, critical thinking, systems dynamic, etc.) (Saviano & Di Nauta, 2011). In this sense, although specific expertise and competences are not only useful but also necessary, being well equipped with the techniques and tools for practicing inter- and transdisciplinarity does not generally resolve the key issue.

Therefore, in our view, to support complex interdisciplinary (as well as transdisciplinary) thinking, specific approaches or techniques of research integration and implementation are useful but a more practical level. We believe that a more *generalizable* 'horizontal' type of knowledge is necessary to integrate, in turn, practical approaches and guide their use. Our idea, in a certain sense, recalls the "Dynamic Usage of Models" proposed by Minati and Pessa that is based on "the ability to *systematically use* the available models" (Minati & Pessa, 2006: x; see also: Minati & Pessa, 2018). In this ability we see the role of a systems thinking 'intelligence' (Mella, 2012) that guides the use of existing knowledge.

To clarify our arguments about the development of such intelligence, in the following sections, we highlight the distinction between generalizable and case-specific knowledge, reaffirming the fundamental role of systems thinking as a response to the need of sharable horizontal knowledge.

## 2.2 From Case Specific to Generalizable Knowledge: The Iter for Complex Interdisciplinary Thinking

In the study entitled "Physical and biological emergence: Are they different?" in the volume *Systemics of Emergence: Research and Development* (Minati et al., 2006a), Eliano Pessa reasons that it is useful to recall 'abstraction' as an approach to shift from specific to generalizable knowledge, valid to the aims or our reflection.

In the work, Pessa, comparing the features of models of emergence introduced within theoretical physics with the requirements coming from observations of biological self-organization, argues that "notwithstanding the deep differences between biological and non-biological systems, the methods of theoretical physics could, in principle, account even for the main features of biological emergence" (Pessa, 2006: 355). In particular, interestingly to our aims, Pessa wonders: "can we resort to suitable generalizations of models describing physical emergence to account for observed features of biological emergence, or, on the contrary, to deal with the latter we need an entirely new approach, incompatible with the one adopted to model physical emergence?"

Pessa addresses the question by providing a formal demonstration of his claims, verified under several circumstances. It is, of course, outside our aims the 'specific' aspects of the approach used by Pessa, which concern a problem common to physics and biology disciplines, although implicitly recalling more general epistemological aspects; rather, we are interested just in these 'generalizable' aspects.

We believe that systems thinking and related theory still represent one of the most useful frameworks of reference for developing truly generalizable knowledge; the kind of knowledge that is common to any discipline, hence a basis for a shared dialogue. Indeed, systems theory remains one of the most effective attempts to integrate the different claims or subject matter as constructed in separate disciplines (Jansen, 2009). In this sense, overcoming the criticism toward the adoption of systems theory in social sciences, we reaffirm the relevance of systems thinking and systems theory as a common basis for all disciplines, in the same way it was conceived to be when a general systems theory was elaborated by Ludwig von Bertalanffy (1956, 1962, 1968, 1969) and subsequently developed (Buckley, 1967; Rapoport, 1986; Laszlo & Krippner, 1998; Luhmann et al., 2013; Minati et al., 2016). The generalizability of systems thinking provides a way to finding the sought common denominator for bridging disciplines in an interdisciplinary space of knowledge sharing that horizontally cross-cuts the boundaries of interacting disciplines.

Therefore, we reverse here the original criticism against systems theory, when, during the Seventies of the last century, many scholars gradually abandoned their original interest toward systems theory, convinced that it was too abstract to offer a reliable representation of the specificities of phenomena, and proposing contingency views as approaches allowing less abstraction, more explicit patterns of relationships, and more applicable theory (Kast & Rosenzweig, 1975). After decades of developing ever more specific, explicit, applied knowledge, a contrasting problem has emerged: scholars now seem to lack general references for shaping their mindsets, and while their expertise provides excellent solutions to an ever wider range of practical problems, they seem to become ever less able to even comprehend and frame problems, which subsequently appear to them too complex. Believing that complexity is not so much a character of the problem to solve as a condition of the problem solver, clearly, they are the decision makers' interpretative and understanding capabilities that are questioned (Barile 2009a, b ). It seems that the more knowledge developed is specific of defined problematic contexts, the more the capability to look behind and beyond that specificity is weakening.

In our view, what appears is an incapability of decision makers to read the general principles that govern all the specific manifestations of human as well as nature's behaviors, abstracting generalizable knowledge from the various problematic contexts to govern. In this sense, we think that also scientific thinkers, i.e. scholars, as those mainly responsible for the progress of knowledge, need to recover and share a more abstract knowledge, derived from very general principles that govern the world in all its forms, so to reacquire the capability to see its irreducible unity (Capra & Luisi, 2014).

Such intelligence is necessary to develop an interdisciplinary dialogue among disciplines: a common language and shared mindsets that make communication effective allowing reciprocal understanding. In other words, a meta-level knowledge, which can guide the integration of more specific approaches through the searching of their commonalities. For example, complex systems science and sustainability science develop their own models and interpretative schemes useful for facing specific problems; the two sciences, however, are characterized by similar complexity and have both common roots in systems theory (Barile & Saviano, 2018; Saviano et al., 2019). Subsequently, to integrate them for facing complex intertwined problems, generalizable systems thinking knowledge can be abstracted from the specificity of the respective problematic contexts in order to create a shared area of knowledge development.

Pessa provided a clear example of what we intend to highlight: to sustain his interpretative proposal, he resorted to "suitable generalizations" for biological and non-biological systems, which allowed him to use the same theoretical methods to account for emergence in two different scientific fields. In other words, 'generalization' allowed shifting, under certain conditions, from one theoretical context to another.

More interestingly, the way the "suitability" of generalizations is verified by Pessa refers to "two relevant features [ . . . ]: the first describing the role played by general principles, and the second related to the existence of individual differences between the single components of the system under study" (Pessa, 2006: 359). The search for "general principles" and "individual differences" is relevant to distinguish between "Ideal models" and "Non-ideal models", where the attribute "ideal" is used "to denote the model in which evolution laws, transition rules, constraints, boundary conditions, as well as every other circumstance underlying the production of system's behaviors, are nothing but a consequence of general principles" (Ibid.). Accordingly, Non-ideal models refer to "models in which behaviors (local or global) are nothing but a consequence of the introduction of suitable local evolution rules, supplemented by a fortunate choice of initial and boundary conditions, as well as of right parameter values" (Ibid.).

Hence, the focus should be on finding or creating suitable generalizations, laws, transition rules, constraints, etc., which are a consequence of general principles.

Moving among different theoretical fields is, however, still a challenge to address in the creation of new knowledge, leveraging the existing; the goal to pursue is still exploiting existing knowledge to explore its application in new problematic contexts (March, 1991). More in general, as underlined, for addressing complex issues, which are typically multi-dimensional, multi-perspective, multi-stakeholder, hence requires the integration of inter- and transdisciplinary knowledge, whereby the possibility to cross different knowledge domains to integrate their contribution is fundamental.

If we recognize that under observed complex phenomena, common rules of systemic functioning lie, following Pessa's reasoning, it is possible to 'see' the general systems principles (of which such rules are consequences) that underlie the observed phenomena, looking behind and beyond the specificity of the "individual differences" and making "suitable generalizations". This generalizable systems thought, by virtue of its high level of abstraction, can represent the sought common denominator on which to base inter- and transdisciplinary sharing of knowledge for the integration and implementation of multiple knowledge resources.

In the following sections, we illustrate our proposal of a systems thinking reference for inter- and transdisciplinarity by briefly describing key reference elements and schemes of the Viable Systems Approach (*vSA*).

## 3 Systems Thinking as a Common Denominator for Approaching Interdisciplinarity: A Proposal That Stems from *vSA*

Our introductory reasoning about the need of a re-founding of human values and the role of scholars in the process has led us to recognize the opportunity of promoting inter- and transdisciplinary knowledge sharing to face the complexity of decision-making in conditions of uncertainty. In effect, we believe that such collective sharing can provide not only interdisciplinary knowledge useful to deal with complex issues but also a way towards the proposed re-founding of human values.

We illustrate our view, in the next sub-section by means of:

1. The *Viable Systems Approach* (*v*S*A*), as an interpretative and governance methodology for organizations developed within the field of managerial sciences to integrate existing problem solving competences with more general decision making capabilities necessary to address the emerging complexity (Barile, 2000, 2008, 2009a, b; Golinelli, 2000a, b, 2010, 2022; Barile et al., 2011, 2012b, 2014, 2016a, b; Barile & Saviano, 2011).
2. The *Information Variety Model*, as a general scheme of reference developed within *v*S*A* to frame the representation of a system's knowledge structure and dynamic (Barile, 2009a, b).
3. A *v*S*A* revised version of the *T-shaped Model*, developed within the field of Service Science to frame the increasing need of knowledge that integrates vertical expertise in disciplines/systems with horizontal boundary-crossing capabilities, necessary to address complex problems (Guest, 1991; Leonard, 1995; Hansen, 2001; Donofrio et al., 2010; Barile et al., 2012a, 2015; Barile & Saviano, 2013a, b; Demirkan & Spohrer, 2015; Saviano et al., 2017a, b).

## 3.1 The Viable Systems Approach (vSA) as a Systems Thinking Reference to Develop Interdisciplinary Knowledge

As highlighted, during recent decades, while experiencing a growing complexity, decision makers of social and business organizations have had to recognize the increasing inadequacy of current models and interpretative schemes in the comprehending and governing of social, economic and environmental dynamics. The highly specialized tools and models of management, focused on problem solving, developed as an outcome of the vertical specialization of the industrial era, appear increasingly inadequate to address complex decisions. It has long been evident, in fact, "the inadequateness of concepts and language based on industrial knowledge still used in current practices by managers to cope with problems of the post-industrial societies characterised by non-linear process of emergence and acquisition of properties" (Minati, 2012: 350).

In this context, a stream of management scholars started a deep rethinking of consolidated management models, recognizing the need to recover more powerful interpretative schemes. By building on the key principles of the von Bertalanffy's General Systems Theory (von Bertallanfy, 1962, 1968, 1969) and on an updated version of Beer's Viable System Model (1972; Espejo, 1990; Espejo & Reyes, 2011), the Viable Systems Approach (*v*S*A*) developed a unitary set of general interpretative schemes as a reference for studying and governing social and, in particular, business organizations on the basis of five founding principles[3] (Barile, 2009a, b; Barile et al., 2015).

Starting from the early works of Sergio Barile (2000, 2008, 2009a, b) and Gaetano Golinelli (2000a, b, 2005, 2010), *vSa* has been developed over recent decades providing generalizable models of individual and organizations viewed as viable systems (Espejo & Harnden, 1989; Espejo & Gill, 1997; Yolles, 1999). In Table 1, ten main groups of concepts and relative propositions are listed to summarize key constructs of *vSa*.

Through 'suitable generalizations', *vSa* enables the reading of behavioral rules underlying observed phenomena that are common to different problematic contexts and fields. By overcoming the limits of a still dominant analytical-reductionist approach which tends to focus on micro aspects that are specific to observed phenomena, the systems approach helps to integrate different perspectives within a unitary framework of generalizable knowledge. Therefore, a systems approach to interdisciplinarity should be intended as a way to open up "space for thinking in a non-reductionist way about multiple determinations without rejecting the value of single disciplines" (Jansen, 2009: 172).

On these bases, *vSa* has developed multiple research trajectories both horizontal (enrichment of the paradigm by deepening general level issues such as complexity, sustainability, etc.) and vertical (application of *vSa* general interpretative schemes to specific problematic contexts). In this sense, the systems approach allows combining research by following both horizontal (dynamic capabilities) and vertical (competences) trajectories in line with the view of "T-Shaped" knowledge (Barile & Saviano 2013a, b; Barile et al., 2014).

The distinction between horizontal and vertical knowledge is useful to interpret the process of generalization, as will be seen hereafter.

## *3.2 The Information Variety Model as a General Reference for Representing a System's Knowledge Variety*

Among the main general schemes of *vSa*, listed in Table 1, Information Variety (Barile, 2009a, b) represents one of the most useful advancements of *vSa*.

---

[3] The five principles of *VSA* (Barile et al., 2015): 1. *Survival*: A viable system, inserted in a specific context, has the primary purpose of survival. 2. *Eidos*: The viable system, in its ontological qualification, can be conceived from a double perspective: that of the structure and that of the system. 3. *Ethos*: The viable system, in its behavioral qualification, is characterized by two logically distinct areas: that of deciding and that of acting. 4. *Isotropy*: The viable system, in its existential dynamics, is directed towards the pursuit of strategies and the achievement of objectives in a daily life characterized by constant interaction with suprasystems and subsystems from which and to which, respectively, it draws and provides guidelines and rules. 5. *Exhaustiveness*: For a viable system, all entities external to itself or are also viable systems, or are components of higher-level viable systems.

**Table 1** Key concepts and propositions derived from the five *VSA* principles

| Key concepts | Propositions |
|---|---|
| Survival | A viable system aims at surviving in its context. |
| Context | The viable system's context is made up of other viable systems. |
| Viability and sustainability | A system remains viable in its context if it is able to establish (sustainable) relationships with other systems to gain access to resources necessary for its own functioning. |
| Reductionism and holism | A system can be studied by focusing both on its parts as components of the structure and on its whole functioning directed at the achievement of a goal. |
| Structure-system perspective | Any viable system can be observed through a dual perspective: that of the structure and that of the system. The system emerges from the structure when relations are finalized to the achievement of a goal. |
| Boundaries | A viable system has no boundaries. Only structural borders can be drawn while the systemic configuration dynamically changes depending on the changing relationship it develops in the context in the pursuing of its goals. |
| Complexity | Complexity does not objectively characterize a phenomenon but subjectively arises when the decision maker is not able to comprehend experienced phenomena through consolidated interpretative schemes. |
| Information variety | A viable system can be viewed as an information variety that qualifies its cognitive endowment in terms of information units, interpretative schemes, and categorical values. |
| Deciding and acting | Viable systems are characterized by deciding and acting functions. |
| Consonance and resonance | A viable system develops conditions of harmonic relationship with (supra)systems (consonance) in order to create value through synergistic interaction (resonance). Consonance can be diversely achieved by acting on the three dimensions of the information variety. |

Source: Barile et Saviano: www.asvsa.org

By integrating cognitivist and constructivist theories (Weick, 1979; Papert, 1986; Meyrowitz, 1995; Hatch, 1999), *VSA* proposes an interpretation of the viable systems' knowledge processes that is based on their representation as Information Varieties (Barile, 2009a). In this respect, the Information Variety Model (IVM) frames a three-dimensional representation in which the viable systems' knowledge is viewed as articulated in *information units*, *interpretative schemes* and *categorical values*. Such dimensions, however, are not intended to express 'proportions' of information variety as in typical spatial representations of material entities; in fact, they are not structural but 'systemic' dimensions, whose meaning depends on the subjective perspective of the observer and on the specific context of reference (Barile et al., 2012a). In the metaphorical representation of Fig. 1, Information Variety is conceived as an atom in which the information units are the electrons, and the interpretative schemes and categorical values are the nucleus: the information units are 'attracted' by the nucleus and orbit around it at various distances.

The metaphorical recalling of the principles of physics into a context of study of organizations allows the transposition of knowledge from the one field to the other.

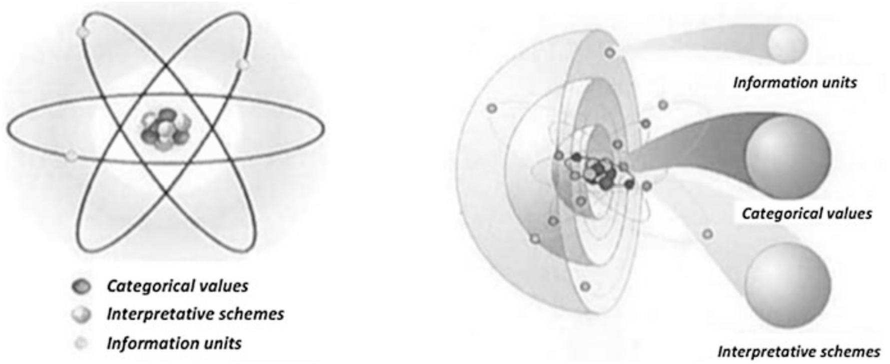


**Fig. 1** A metaphorical representation of the Information Variety Model (IVM). (Source: Barile, 2009a, b: 88. www.asvsa.org)

Here, however, we limit ourselves to provide essential elements useful to our aims, starting from the definition of the three dimension of the model.

The information units are conceived as the 'structural' composition of knowledge, which is the specification of the total amount of data held by the viable system including all that it can perceive or can further determine by processing and transforming the data into information significant to the knowledge process.

The interpretation schemes represent the knowledge patterns through which information is organized within the viable system's whole variety. The concept of 'scheme' is that of an organizing structure of past and current experience (Penna & Pessa, 1994). Without such logical interpretative schemes, every piece of information would appear to us as new every time we perceive it, and consequently, we would need to create a new interpretation model to explain and comprehend it every time. It is through these interpretation schemes that we transform generic data into contextualized information (Barile et al., 2012a; De Toni & Comello, 2005). vSA distinguishes two kinds of interpretation schemes: general and specific. Whereas the former is compressed and potentially active, the latter is 'in use'; in other words, general schemes become specific schemes when they are used in a specific context, enabling a generation of new knowledge through a process of contextualization.

It is worth noting here that the reverse process of contextualization is abstraction: general schemes are what we find when we abstract knowledge from the specific problematic context on the basis of suitable generalizations that allow us to see the principles of which the observed specificities are consequences. This is a key process in an interdisciplinary knowledge co-creation context in which the knowledge to integrate is disciplinary-specific. Searching for common rules of systemic functioning should lead to sharable interpretative schemes that facilitate interaction among different disciplines.

At the next level of IVM, an aspect is highlighted that is less usually included in the practice of systems thinking, and which is a peculiar contribution of vSA: the role of values. The so-called categorical values represent, in fact, the most relevant dimension of the Information Variety. They characterize the viable system's deep-rooted strong beliefs that over time define the system's unique identity. Categorical values are responsible for accepting or refusing rational elaborations and for directing the functioning of the interpretative schemes. They are strictly connected to the emotional level of the interaction process. The role of the categorical values sheds light on the core of our reflection: we believe, in fact, that interaction in a multidisciplinary context, not limited to the exchange of information units but extended to the sharing and integration of interpretative schemes, thanks to an abstraction process that leverages the common systems thinking roots, not only facilitate reciprocal understanding, hence the sharing of knowledge, but also the alignment of values.

According to IVM, in fact, the process and outcomes of interaction between individuals or organizations as viable systems depend on their conditions of *consonance* and *resonance*: the former qualifies a structural compatibility as a necessary (but not sufficient) condition to render the intra- and inter-systemic interactions possible and potentially effective; the latter qualifies the degree of 'alignment' between the interacting varieties determined by the degree of consonance. As will be highlighted, the concepts of consonance and resonance can be useful to analyse a knowledge co-creation context such as that of interdisciplinarity.

In actual fact, interdisciplinarity has been studied mainly to comprehend whether and how it can be more effective and successful compared to disciplinary knowledge, often questioning the latter's assumed superiority; in their critical assessment of interdisciplinarity, Jacobs and Frickel (2009) call for more research on various aspects and, interestingly, conclude by arguing that "many topics in this area require serious conceptual advances as well as creative collection of new data. For example, can general criteria be developed that would indicate the appropriate level of communication between disciplines? Can general criteria be developed for the evaluation of interdisciplinary research?" (Jacobs & Frickel, 2009: 61). The development of "general criteria" recalls the need to recover a more general level knowledge useful to govern the integration of the more specific.

In line with this view, in interdisciplinary interaction contexts, focus should be placed on the sharing of not only knowledge from facts, i.e. data, information, solutions etc., but also the approaches used to develop and interpret such information, behind which there are general interpretative schemes used to process incoming information.

Therefore, the abstraction process suggested by *vSA* leads bottom up from the information units to general scheme sharing and creates a context for categorical values to act determining the conditions of consonance and the result of resonance in the interdisciplinary co-creation of knowledge (Wieland et al., 2012; Barile & Saviano, 2013a, b; Polese et al., 2017, 2018; Polese, 2018).

A synthesis of our interpretative proposal will be discussed and illustrated in the following section using the T-shaped Model.

### *3.3 The T-Shaped Model as a General Reference for Developing Horizontal Interdisciplinary Knowledge*

The Information Variety Model (IVM) provides a general reference for discussing interdisciplinarity in knowledge co-creation contexts. With respect to the aims of this contribution, through IVM it is possible to highlight that:

1. Although disciplinary knowledge continues to be extremely important for progress, there is increasingly the need of multi-, inter- and trans-disciplinary approaches to co-create knowledge.
2. Interdisciplinarity requires co-creation of knowledge through sharing of not only information units but also interpretative schemes and, above all, values.
3. Interdisciplinarity practised through systems thinking can become an iter for a values alignment that is necessary for the re-founding we are seeking of a new paradigm.

As highlighted, cross-cutting disciplinary boundaries for inter- but also transdisciplinarity is no new problem in science (Klein, 1990). Despite numerous successful experiences, interaction between sciences as well as disciplines is still practically and also conceptually difficult; in fact, interdisciplinary institutes and research groups struggle to survive and often fail (Lélé & Norgaard, 2005). Traditional obstacles to interdisciplinarity were first of all of an institutional and even political type, as "scientific disciplines constitute the *modern social order* of knowledge" (Lélé & Norgaard, 2005: xi), which is difficult to change. On the other hand, it has been recognized that many interdisciplinary projects do not per se make very relevant gains while established academic disciplines show themselves to be ever more dynamic centres of knowledge production "open to external developments even while insisting on internal standards" (Jacobs & Frickel, 2009: 44). Moreover, when interdisciplinary research is successful, it can tend to become established giving rise to new fields of inquiry, from which new rounds of differentiation and fragmentation can start. Somehow, interdisciplinarity itself can be viewed from a 'disciplinary' perspective.

Actually, to make interdisciplinarity successful not so relevant are the contents of knowledge but rather the methods to develop them. As underlined, the kind of knowledge required must support boundary crossing interaction, allowing effective communication and reciprocal understanding between interacting disciplines so to develop new knowledge starting from existing one.

Going up from specific and contextualized to more general knowledge, common to various fields, is a way to identify the sought common denominator useful to approach interdisciplinarity. Disciplinary knowledge, developed over time within the 'borders' of the scientific communities, is ever more specifically conceived to increase humans' well-being and the quality of every day life; in this sense, it follows a 'vertical' pathway. Conversely, interdisciplinary boundary-crossing knowledge allows horizontally connecting multiple vertical knowledge expertise.

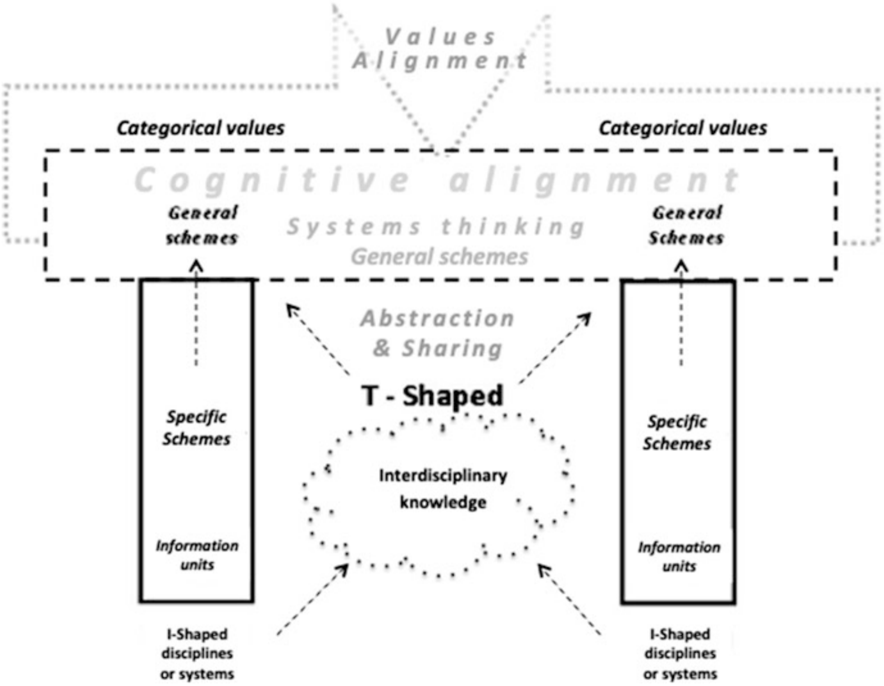


**Fig. 2** A *VSA* T-shaped Model of interdisciplinarity. (Source: Elaboration from Barile et al.,. 2012b. www.asvsa.org)

As discussed, the definition of horizontal knowledge can be based on the search of generalizable commonalties that characterize the scientific fields to integrate. One way is to recognize the common systemic behaviour of the various phenomena investigated. Hence, this systemic nature can represent the sought generalizable underlying principle, and systems thinking the sought common denominator that provides models and tools useful to bridge different disciplinary knowledge resources,

The way multiple vertical knowledge can be bridged, in our view, using systems thinking and specifically the *VSA* general schemes is outlined in Fig. 2 in which the Information Variety Model is used as an explicative tool to develop the horizontal bar of a T-shaped Model of knowledge (Barile et al., 2012b).

In Fig. 2 a simplified representation is proposed of a potential dynamic of interaction between different disciplines/systems and the possible interdisciplinary outcome. As illustrated, effective interaction is at the same time made possible by consonant categorical values and reinforced by the values alignment outcome. It is in this sense that we believe that interdisciplinarity is, on the one hand, facilitated by conditions of consonance between interacting entities, by virtue of the positive role of converging (or at least compatible) categorical values and, on the other, enhanced by the subsequent effect of consonant interaction, producing what *vSA* calls resonance, an acceleration of consonance.

If such a process occurs, our hypothesis is that it can contribute to the creation of the auspicated conditions for a re-founding of values necessary to a new paradigm. In fact, while reducing the 'cognitive distance' (Nooteboom, 2006) between the multiple interacting entities, interdisciplinary practice creates a context that can favor a progressive alignment. Moreover, if we agree that the more "distant" the interacting entities are, the greater is the potential of generating new knowledge making effective interdisciplinarity, the threats connected to the initial distance, that is to the extent that cognition differs, which make difficult reciprocal understanding and collaboration (Nooteboom, 2006), can be attenuated by basing interaction on elements of convergence, which should happen using systems thinking as a common denominator. In this respect, Nooteboom's reasoning appears to support our view. As the Scholar affirms (2006: 14): "collaboration across (greater or smaller) cognitive distance forces one to try and apply one's knowledge in a novel context, in this case the practice of the partner (generalization) process now is reciprocal. Partners can help each other in fitting in elements from their practice into hybridization of the partner's practice. Next, they can try to jointly find novel design principles for a synthesis, in a new form."

In short, on the basis of IVM, we can assume that interaction between different knowledge varieties (disciplines/systems) can result in a variation of the initial variety whose degree depends on the initial and progressive conditions of consonance. When different but consonant 'I-shaped' knowledge varieties (vertical expertise) interact, a cross-fertilization potential can emerge leading to a "hybridization" of knowledge using sharable interpretative schemes, like those offered by systems thinking schemes. In this sense, systems thinking acts as a bridging meta-knowledge.

The envisioned abstraction up to the level of general schemes and even beyond to the dimension of values that lie behind decisions, choices and behaviors of humans' individual and organized systems, leads to a dimension that, metaphorically speaking, is somehow 'absolute' because, at that level, the interacting 'parts' vanish and become 'indistinct' within the whole, by virtue of a deep sharing. In effect, information is 'compressed' in patterns, which can be traced to general schemes, which in turn can be traced to categorical values.

In order to imagine how the different 'dimensions' of the knowledge variety should be intended to act in a sharing context, in Fig. 3 we provide a potential evolutionary representation of the *vSA* T-shaped Model of knowledge in which the horizontal bar becomes a 'sphere' (e.g. a three-dimensional representation recalling the representation of Fig. 1), in which it is compacted around a central 'nucleus' whose core is made up of deep-rooted values. From the 'mass' of such nucleus an extremely wide range of knowledge potentialities arises thanks to the rich endowment of general schemes that can support the development of contextualized or case specific knowledge in many different contexts.

Of course, a similar representation may appear a 'too abstract' reinterpretation of the T-shaped idea, compared to its common versions. Actually, we exalt and raise

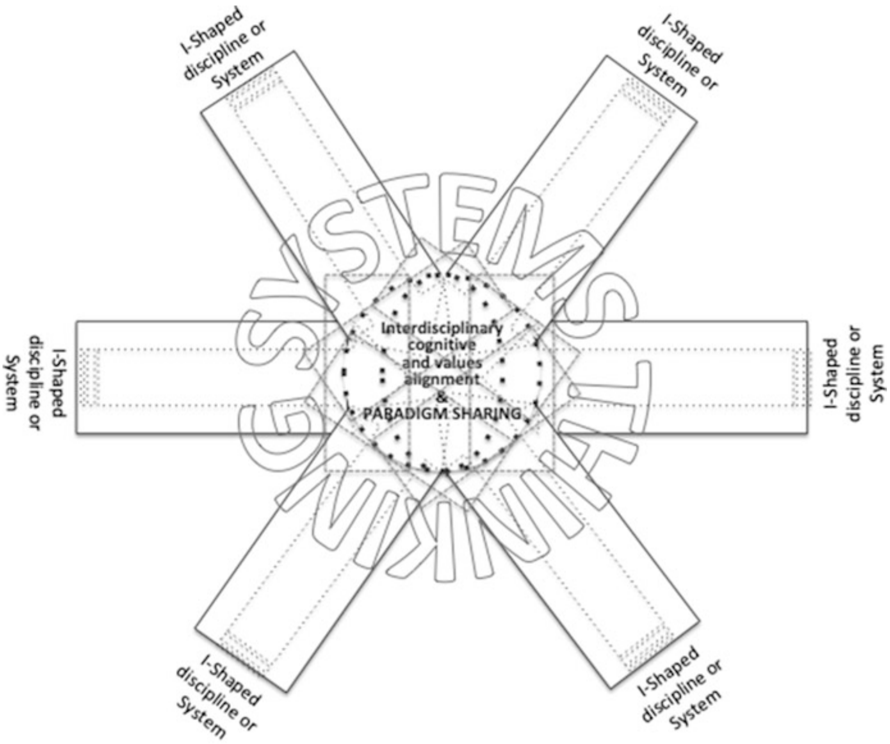


**Fig. 3** A *VSA* T-shaped Model of knowledge sharing. (Source: Saviano et Barile. www.asvsa.org)

the idea of horizontality to the extent that, at a very high and general level, it is made of a few very general schemes with a high knowledge potential.

The ‘I’, ‘T’ and ‘spherical’ interpretation of knowledge variety can be useful also to distinguish different knowledge profiles, from informed, to smart, up to wise people, as will be highlight in the concluding section of our reflection.

## 4 Concluding Remarks

In a knowledge society fast changing like the one we are living in now, whose unpredictable evolution is posing many among the most critical challenges to humans, as highlighted in the introductory premise, it seems that the kind of knowledge most required is what we would call ‘smart knowledge’, characterized by high technological skills and the use of advanced tools, such as Artificial Intelligence (AI) (McCarthy, 2007). The role played by values in the knowledge exchange process we auspicate, however, would remain highly humans-centered, as highlighted by the Information Variety Model and also underlined in the T-shaped version of the same.

The query we posit is: how will and should the values dimension evolve in the future? This is where our preoccupation about the future mainly lies as it seems that the space for values is progressively reducing. Values, however, are those that make people, as well as organized systems, wise, not only smart (Herczeg, 2010; Carayannis et al., 2018; Barile et al., 2018a, b, 2020).

Therefore, in our view, a re-founding of values necessary to a new paradigm makes the abstraction process discussed above even more critical in that it allows knowledge be characterized not only by technological elements but also human ones. And human intelligence, compared to artificial, will always bring into knowledge an irreducible values dimension (Mella, 2012).

Accordingly, we believe that intense human-centered knowledge exchange interaction, such as that characterizing the role of scholars, can be the key for the next paradigm shift, as to a certain extent also envisioned in the view of the emerging Society 5.0 e.g. as a platform society (Van Dijck et al., 2018) in which the digitized world should once again have humans at the core (Fukuyama, 2018).

Therefore, creating and developing conditions of consonance between interacting knowledge varieties, such as scientific disciplines, can make the practice of interdisciplinarity a way to recovering and recognizing the role of values in any context of decision and choice in which the emerging of a collective consciousness can be the key for the sharing of an innovative paradigm.

**Acknowledgments** The Authors are the extenders of thoughts that come from the extensive sharing of reflections, over the last two decades, within the scientific community of the *Viable Systems Approach* (*VSA*), whose development has been possible only by virtue of fundamental readings of the works of many among the most impactful thinkers including Ashby (1958), Atlan (1987), Beer (1981, 1985), Bogdanov (1913), Buckley (1968), Capra (1997), Clark (1993), De Toni (2005), Emery & Trist (1965), Heisenberg (1971), Korzybski (1933), Lovelock (1979), Maturana & Varela (1975), Mella (2012), Minati et al. (2006a, b), Minsky (1974), Morin (2005), Pessa (2002), Popper (1996), Saraceno (1969), Simon (1990), Vernadskij (1986), von Bertalanffy (1956), von Foerster (1984), Wiener (1948), to name but a few.

## References

Ashby, W. R. (1958). General systems theory as a new discipline. *General Systems (Yearbook of the Society for the Advancement of General Systems Theory), 3*, 1–6.

Ashby, W. (1968). Principles of self-organizing system. In W. Buckeley (Ed.), *Systems research of behavioral science* (pp. 116–117). Aldine Transaction.

Atlan, H. (1987). Self creation of meaning. *Physica Scripta, 36*(3), 563.

Badinelli, R., Barile, S., Ng, I., Polese, F., Saviano, M., & Di Nauta, P. (2012). Viable service systems and decision making in service management. *Journal of Service Management, 23*(4), 497–526.

Bammer, G. (2017). Should we discipline interdisciplinarity? *Palgrave Communications, 3*(1), 1–4.

Bammer, G., O'Rourke, M., O'Connell, D., Neuhauser, L., Midgley, G., Klein, J. T., et al. (2020). Expertise in research integration and implementation for tackling complex problems: When is it needed, where can it be found and how can it be strengthened? *Palgrave Communications, 6*(1), 1–16.

Barile, S. (Ed.). (2000). *Contributi sul pensiero sistemico in economia d'impresa*. Arnia.

Barile, S. (2008). *L'impresa come sistema* (2nd ed.). Giappichelli.

Barile, S. (2009a). *Management sistemico vitale*. Giappichelli.

Barile, S. (2009b). The dynamic of informative varieties in the processes of decision making. In *The 3rd International Conference on Knowledge Generation, Communication and Management KGCM*, Orlando, Florida.

Barile, S., Bassano, C., Calabrese, M., Confetto, M. G., Di Nauta, P., Piciocchi, P., et al. (2011). *Contributions to theoretical and practical advances in management. A viable systems approach (VSA)* (pp. 1–288). International Printing.

Barile, S., Bassano, C., Lettieri, M., Piciocchi, P., & Saviano, M. (2020, July). Intelligence augmentation (IA) in complex decision making: A new view of the vSa concept of relevance. In *International Conference on Applied Human Factors and Ergonomics* (pp. 251–258). Springer.

Barile, S., Di Nauta, P., & Iandolo, F. (2016a). *La decostruzione della complessità. Studi MOA—Collana di Management e Organizzazione Aziendale*. Editrice Minerva Bancaria.

Barile, S., Espejo, R., Perko, I., & Saviano, M. (Eds.). (2018a). *Cybernetics and systems* (pp. 297–304). Routledge.

Barile, S., Franco, G., Nota, G., & Saviano, M. (2012a). Structure and dynamics of a "T-Shaped" knowledge: From individuals to cooperating communities of practice. *Service Science, 4*(2), 161–180.

Barile, S., Lusch, R., Reynoso, J., Saviano, M., & Spohrer, J. (2016b). Systems, networks, and ecosystems in service research. *Journal of Service Management, 27*(4), 652–674.

Barile, S., Pels, J., Polese, F., & Saviano, M. (2012b). An introduction to the viable systems approach and its contribution to marketing. *Journal of Business Market Management, 5*(2), 54–78.

Barile, S., Piciocchi, P., Bassano, C., Spohrer, J., & Pietronudo, M. C. (2018b, July). Re-defining the role of artificial intelligence (AI) in wiser service systems. In *International Conference on Applied Human Factors and Ergonomics* (pp. 159–170). Springer.

Barile, S., & Saviano, M. (2011). Foundations of systems thinking: The structure-system paradigm. In *Various authors, contributions to theoretical and practical advances in management. A viable systems approach (VSA)* (pp. 1–24). ASVSA, Associazione per la Ricerca sui Sistemi Vitali. International Printing.

Barile, S., & Saviano, M. (2013a). An introduction to a value co-creation model, viability, syntropy and resonance in dyadic interaction. *Syntropy, 2*(2), 69–89.

Barile, S., & Saviano, M. (2013b). Dynamic capabilities and T-shaped knowledge: A viable systems approach. In *Contributions to theoretical and practical advances in management. A viable systems approach (VSA)*. Aracne.

Barile, S., & Saviano, M. (2018). Complexity and sustainability in management: Insights from a systems perspective. In *Social dynamics in a systems perspective* (pp. 39–63). Springer.

Barile, S., Saviano, M., Iandolo, F., & Calabrese, M. (2014). The viable systems approach and its contribution to the analysis of sustainable business behaviors. *Systems Research and Behavioral Science, 31*(6), 683–695.

Barile, S., Saviano, M., & Simone, C. (2015). Service economy, knowledge, and the need for T-shaped innovators. *World Wide Web, 18*(4), 1177–1197.

Beer, S. (1981). *Brain of the firm: The managerial cybernetics of organization*. Wiley.

Beer, S. (1985). *Diagnosing the system for organization*. Wiley.

Bogdanov, A. (1913). *Saggi di scienza dell'organizzazione*. Theoria, Napoli.

Brewer, G. D. (1999). The challenges of interdisciplinarity. *Policy Sciences, 32*(4), 327–337.

Buckley, W. (1967). *Sociology and modern systems theory*. Prentice-Hall.

Buckley, W. (1968). *Systems research of behavioral science*. Aldine Transaction.

Capra, F. (1997). *The web of life*. Flamingo.

Capra, F., & Luisi, P. L. (2014). *The systems view of life: A unifying vision*. Cambridge University Press.

Carayannis, E. G., Del Giudice, M., Saviano, M., & Caputo, F. (2018). Beyond big data: From smart to wise knowledge management. In S. Barile, R. Espejo, I. Perko, & M. Saviano (Eds.), *Cybernetics and systems* (pp. 297–304). Routledge.

Clark, A. (1993). *Associative engines*. MIT Press.

De Toni, A. F., & Comello, L. (2005). *Prede o ragni*. Utet.

Demirkan, H., & Spohrer, J. (2015). T-shaped innovators: Identifying the right talent to support service innovation. *Research-Technology Management, 58*(5), 12–15.

Donofrio, N., Spohrer, J., Zadeh, H. S., & Demirkan, H. (2010). Driven medical education and practice: A case for T-shaped professionals. *MJA Viewpoint*.

Dube, B. (2021). Why cross and mix disciplines and methodologies?: Multiple meanings of interdisciplinarity and pluralism in ecological economics. *Ecological Economics, 179*, 106827.

Emery, F. E., & Trist, E. L. (1965). The casual texture of organizational environments. *Human Relations, 18*, 21–32.

Espejo, R. (1990). The viable system model. *Systems Practice, 3*(3), 219–221.

Espejo, R., & Gill, A. (1997). *The viable system model as a framework for understanding organizations*. Phrontis Limited/SYNCHO Limited.

Espejo, R., & Harnden, R. (1989). The VSM: An ongoing conversation. In *The viable system model: Interpretations and applications of Stafford Beer's VSM*. John Wiley and Sons Ltd.

Espejo, R., & Reyes, A. (2011). *Organizational systems: Managing complexity with the viable system model*. Springer Science & Business Media.

Fadeeva, Z., Mochizuki, Y., & Parker, J. (2010). Competencies for interdisciplinarity in higher education. *International Journal of Sustainability in Higher Education, 514*, 2010.

Frodeman, R., Klein, J. T., & Pacheco, R. C. D. S. (Eds.). (2017). *The Oxford handbook of interdisciplinarity*. Oxford University Press.

Fukuyama, M. (2018). Society 5.0: Aiming for a new human-centered society. *Japan Spotlight, 27*, 47–50.

Golinelli, G. M. (2000a). *L'approccio sistemico al governo dell'impresa* (Vol. I). Cedam.

Golinelli, G. M. (2000b). *L'approccio sistemico al governo dell'impresa* (Vol. II). Cedam.

Golinelli, G. M. (2010). *Viable systems approach. Governing business dynamics*. Cedam/Kluwer.

Golinelli, G. M. (2011). *L'Approccio Sistemico Vitale (ASV) al governo dell'impresa*. Cedam.

Guest, D. (1991). The hunt is on for the Renaissance Man of computing. *The Independent, 17*(9).

Hadorn, G. H., Biber-Klemm, S., Grossenbacher-Mansuy, W., Hoffmann-Riem, H., Joye, D., Pohl, C., ... & Zemp, E. (2008). The emergence of ransdisciplinarity as a form of research.In handbook of transdisciplinary research (pp. 19–39). Springer, Dordrecht.

Hansen, M. T. (2001). Introducing T-shaped managers. *Harvard Business Review, 79*(3), 106–116.

Harari, Y. N. (2015). What explains the rise of humans? *TED London*, June 2015. https://www.ted.com/talks/yuval_noah_harari_what_explains_the_rise_of_humans?language=it.

Hatch, M. J. (1999). Organization theory, modern, symbolic-interpretivist, and postmodern perspectives. *Book reviews, 15*(321), 330.

Heisenberg, W. (1971). *Physics and beyond: Encounters and conversation*. Harper and Row.

Herczeg, M. (2010, December). The smart, the intelligent and the wise: Roles and values of interactive technologies. In *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia* (pp. 17–26).

Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology, 35*, 43–65.

Jansen, K. (2009). Implicit sociology, interdisciplinarity and systems theories in agricultural science. *Sociologia Ruralis, 49*(2), 172–188.

Kast, F. E., & Rosenzweig, J. E. (1975). General systems theory: Application for organizations and management. *Academy of Management Journal*, 447–465.

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice*. Wayne State University Press.

Klein, J. T., Wentworth, J., & Sebberson, D. (2001). Interdisciplinarity and the prospect of complexity: The tests of theory. *Issues in Interdisciplinary Studies, 19*, 43–57.

Korzybski, A. (1933). *Science and sanity*. Prentice Hall.

Laszlo, A., & Krippner, S. (1998). Systems theories: Their origins, foundations, and development. *Advances in Psychology-Amsterdam, 126*, 47–76.

Lélé, S., & Norgaard, R. B. (2005). Practicing interdisciplinarity. *BioScience, 55*(11), 967–975.

Leonard, D. (1995). *Wellsprings of knowledge*. Harvard Business School Press.

Lovelock, J. E. (1979). *Gaia, Nuove idee nell'ecologia*. Bollati Boringhieri.

Luhmann, N., Baecker, D., & Gilgen, P. (2013). *Introduction to systems theory*. Polity.

Maturana, H. R., & Varela, F. J. (1975). *Autopoietic systems* (BLC report 9). University of Illinois.

McCarthy, J. (2007). What is artificial intelligence. Computer Science Department, Stanford University, 2.

Mella, P. (2012). *Systems thinking: Intelligence in action* (Vol. 2). Springer Science & Business Media.

Meyrowitz, J. (1995). L'impatto dei media elettronici sul comportamento sociale. In *Oltre il senso del luogo*. Baskerville.

Minati, G. (2012). Knowledge to manage the knowledge society. *The Learning Organization, 9*(4), 350–368.

Minati, G., Abram, M. R., & Pessa, E. (Eds.). (2016). *Towards a post-Bertalanffy systemics*. Springer.

Minati, G., & Pessa, E. (2006). *Collective beings*. Springer Science & Business Media.

Minati, G., & Pessa, E. (2018). *From collective beings to quasi-systems* (pp. 145–185). Springer.

Minati, G., Pessa, E., & Abhram, M. (2006a). *Systemics of emergence: Research and development*. Springer.

Minati, G., Pessa, E., & Abram, M. (Eds.). (2006b). *Systemics of emergence: Research and development*. Springer Science & Business Media.

Minsky, M. (1974). A framework for representing knowledge. *Artificial intelligence* (Memo 306).

Moran, J. (2010). *Interdisciplinarity*. Routledge.

Morin, E. (2005). *Introduction à la pensée complexe*. Seuil.

Newell, W. H. (2007). Decision making in interdisciplinary studies. In G. Morcol (Ed.), *Handbook of decision making*. CRC Press.

Newell, W. H., Wentworth, J., & Sebberson, D. (2001). A theory of interdisciplinary studies. *Issues in Interdisciplinary Studies, 19*, 1–25.

Nooteboom, B. (2006). *Cognitive distance in and between COP's and firms: Where do exploitation and exploration take place, and how are they connected? CentER discussion paper 2007–04*. Tilburg, The Netherlands: Tilburg University.

Papert, S. (1986). *Constructionism: A new opportunity for elementary science education*. Media Laboratory, Epistemology and Learning Group: Massachusetts Institute of Technology.

Penna, M. P., & Pessa, E. (1994). *La rappresentazione della conoscenza*. Armando Editore: Introduzione alla Psicologia dei Processi Cognitivi.

Pessa, E. (2002). What is emergence? In G. Minati & E. Pessa (Eds.), *Emergence in complex cognitive, social and biological systems*. Kluwer.

Pessa, E. (2006). Physical and biological emergence: Are they different? In G. Minati, E. Pessa, & M. Abram (Eds.), *Systemics of emergence: Research and development* (pp. 355–374). Springer.

Polese, F. (2018). Successful value co-creation exchanges: A VSA contribution. In *Social dynamics in a systems perspective* (pp. 19–37). Springer.

Polese, F., Barile, S., Caputo, F., Carrubbo, L., & Waletzky, L. (2018). Determinants for value cocreation and collaborative paths in complex service systems: A focus on (smart) cities. *Service Science, 10*(4), 397–407.

Polese, F., Mele, C., & Gummesson, E. (2017). Value co-creation as a complex adaptive process. *Journal of Service Theory and Practice, 27*(5), 926–929.

Popper, K. R. (1996). *Tutta la vita è un risolvere problemi. Scritti sulla conoscenza, la storia e la politica*. Rusconi.

Rapoport, A. (1986). *General system theory: Essential concepts & applications* (Vol. 10). CRC Press.

Saraceno, P. (1969). L'analisi dei sistemi nella condotta delle imprese. *AA. VV., Scritti in onore di Giordano dell'Amore*, 2.

Saviano, M., Barile, S., Farioli, F., & Orecchini, F. (2019). Strengthening the science–policy–industry interface for progressing toward sustainability: A systems thinking view. *Sustainability Science, 14*(6), 1549–1564.

Saviano, M., Barile, S., Spohrer, J. C., & Caputo, F. (2017a). A service research contribution to the global challenge of sustainability. *Journal of Service Theory and Practice, 27*(5), 951–976.

Saviano, M., & Di Nauta, P. (2011, June). Project management as a compass in complex decision making contexts: A viable systems approach. In D. Caivano, M. T. Baldassarre, F. Garcia, M. Benero, E. Mendes, P. Runeson, A. Sislitti, G. H. Travassos, & G. Visaggio (Eds.), *12th International Conference on Product Focused Software Development and Process Improvement* (pp. 112–119). ACM Association for Computing Machinery.

Saviano, M., Polese, F., Caputo, F., & Walletzký, L. (2017b). The contribution of systems and service research to rethinking higher education programs: A T-shaped model. *Sinergie Italian Journal of Management, 35*, 51–70.

Simon, H. A. (1990). Bounded rationality. In *Utility and probability* (pp. 15–18). Palgrave Macmillan.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Vernadskij, V. I. (1986). *The biosphere*. New York: Synergetic Press.

von Bertalanffy, L. (1956). General system theory. *General Systems, 1*(1), 11–17.

von Bertallanfy, L. (1962). General system theory—A critical review. In W. Buckeley (Ed.), *Systems research of behavioral science* (pp. 1–20). Aldine Transaction.

von Bertallanfy, L. (1968). *General system theory. Foundations, development, applications*. Penguin University Books.

von Bertallanfy, L. (1969). Evolution. Chance or law. In A. Koestler & J. R. Smithies (Eds.), *Beyond reductionism* (pp. 59–84). Hutchinson.

von Foerster, H. (1984). *Observing systems*. Intersystems Publications.

Wieland, H., Polese, F., Vargo, S. L., & Lusch, R. F. (2012). Toward a service (eco)systems perspective on value creation. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), 3*(3), 12–25.

Wiener, N. (1948). *Cybernetics*. MIT Press.

Weick, K. E. (1979). The social psychology of organizing, II Ed., McGraw-Hill.

Yolles, M. (1999). *Management systems: A viable approach*. London: Financial Times Pitman Publishing.

## *Websites*

https://en.wikipedia.org/wiki/Viable_systems_approach
www.asvsa.org
www.ted.com

# Autonomous Systems and the Place of Biology Among Sciences. Perspectives for an Epistemology of Complex Systems

**Leonardo Bich**

**Abstract** This paper discusses the epistemic status of biology from the standpoint of the systemic approach to living systems based on the notion of biological autonomy. This approach aims to provide an understanding of the distinctive character of biological systems and this paper analyses its theoretical and epistemological dimensions. The paper argues that, considered from this perspective, biological systems are examples of emergent phenomena, that the biological domain exhibits special features with respect to other domains, and that biology as a discipline employs some core concepts, such as teleology, function, regulation among others, that are irreducible to those employed in physics and chemistry. It addresses the claim made by Jacques Monod that biology as a science is marginal. It argues that biology is general insofar as it constitutes a paradigmatic example of complexity science, both in terms of how it defines the theoretical object of study and of the epistemology and heuristics employed. As such, biology may provide lessons that can be applied more widely to develop an epistemology of complex systems.

**Keywords** Biological autonomy · Complex systems · Decomposition · Emergence · Explanation · Function · Integration · Organization · Teleology

## 1 Introduction

The question “what is life?”, the title of the seminal essay by Erwin Schrödinger (1944), keeps raising several theoretical and epistemological issues. Theoretical issues concern for example what types of systems living organisms are, the identification of their distinctive features and their differences with respect to other natural systems such as physical, chemical ones, or ecological and social ones, not to mention hard questions such as how life originated. An example of the complexity

---

L. Bich (✉)
Department of Philosophy, IAS-Research Centre for Life, Mind and Society, University of the Basque Country (UPV/EHU), Donostia-San Sebastian, Spain
e-mail: leonardo.bich@ehu.es

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_4

of this question is the wide and intense debate on the definitions of life, characterized by a lack of consensus and by the proliferation of definitions proposed (Popa, 2004; Bich & Green, 2018).

Moreover, the question about life has important epistemological implications. It raises the problem of how to characterize, describe and explain living systems and biological phenomena in general. The study of living systems has shown the inadequacies of the framework of deductive nomological explanation that had dominated discussions of scientific explanation in the middle decades of the twentieth century. The deductive nomological framework had emphasized the importance of scientific laws, with physics as the science of reference. The study of biological systems, instead, has given rise to an interest in notions such as emergence, due to the difficulties or impossibilities of using one level such as that of physical systems and its laws, to account for a different one, the biological, or to understand a system on the basis of a description of its parts alone. In the last decades an increasing attention has been paid to the specific types of explanations used in biology. Recognizing that biologists seldom refer to laws when advancing explanations, Bechtel and Richardson (1993/2010) focused on the fact that biologists often explain a phenomenon by describing the responsible mechanism, identifying the parts involved, their operations, and their organization rather than identifying or referring to laws. This view has been at the origin of the neo mechanistic approach in philosophy of science (see also Machamer et al., 2000; Bechtel & Abrahamsen, 2005; Craver & Darden, 2013; Glennan, 2017).

The distinctiveness of biology and the complexity underlying the notion of life may lead one to inquire into the epistemic status of biology. Among others, Jacques Monod directly addressed this issue at the beginning of the preface to his book “Chance and Necessity” (Monod, 1970). He did so by somehow acknowledging the ‘special’ character of biological system, and by identifying ‘special’ with ‘rare’. Biological systems constitute only a minimal fraction of natural systems, and therefore, according to Monod, the study of biology might not lead to the discovery of general laws applicable outside the biosphere. In his view biology is marginal.

In this paper I address this issue and question Monod’s claim. I do so by starting from a characterization of living systems in terms of autonomy (Sect. 2), to discuss the general theoretical and epistemological implications of understanding biology as (in a specific sense) special (Sect. 3). As suggested by Robert Rosen “perhaps the first lesson to be learned from biology is that there are lessons to be learned from biology” (2000: 275). I argue (Sect. 4) that biology is not marginal and provides important lessons that may be applied more widely to develop an epistemology of complex systems.

## 2 What Is Life? An Organizational View

When Erwin Schrödinger, as a physicist, addressed the basic question of biology, “what is life?” (Schrödinger, 1944), he brought forth the idea that understanding biological systems might have required extending the framework of physics by providing new laws and concepts. This idea of developing a new physics and introducing new laws was later pursued with a different approach by Stuart Kauffman (2000), among others. This attitude is different from Monod's. Acknowledging, as Schrödinger and Kauffman did, that life cannot be accounted for by current physics, leads to questioning the generality of physics and to expanding or renewing physics itself.

Schrödinger tried to identify what is the invariant element that might allow understanding the distinctive character of biological systems. He identified it with a specific type of order. In opposition with the statistic order of physical systems ("order from disorder"), that might give rise to macroscopic patterns, he identified the invariant element of living systems in what he called "order from order". The basic idea is that of a positional order, embedded in a specific rigid molecular structure, an order which is then propagated throughout the system. Positional order is realized in terms of rigid spatial organization, found at the molecular level in an aperiodic crystal and embedded in a specific sequence (see Bich & Damiano, 2008 for an analysis of different types of order and organizations). In Schrödinger's view, living processes are controlled by highly stable groups of atoms which transmit their structural order to other molecular structures: in contemporary language, from the sequence of bases in the DNA to the sequence of amino acids in proteins. The constancy of positions and sequences and the regularity of the relationships between parts (solid bodies whose form is maintained constant) is what allows organisms and machines such as mechanical clocks to function. Both are characterized as mechanisms and understood in terms of the relative positions of stable (ordered) components, a form of organization.

This is not the only way to look at living systems in terms of organization. Organization can be understood in terms of static spatial relationships such as in this case, but also in terms of dynamic relationships between components and processes that undergo continuous transformations. In fact, if one focuses on processes and on the activities of components that realize living organisms, regularity and constancy might be the exception rather than the rule, or even the sign of a pathology (Bich et al., 2020). Organisms are adaptive systems whose internal dynamics and the fate and behavior of parts depend on the state of the system and its environment (Bich et al., 2016). Living systems constantly modify their components, and their internal mechanisms are highly dynamic. They are continuously modulated, activated or inhibited by regulatory mechanisms. From this point of view, which considers living systems as dynamical, the invariant aspect cannot be found in some structural regularity of sequence or positions at the level of parts but in a relational property at the level of the organization of the system. It is a property of the whole living system. It cannot be referred to any specific component of it; rather, it rests on the peculiar and distinctive way the components—and the processes they are involved in—are related.

A theoretical approach focused on this type of organization has been developed at the crossroad between cybernetics and systems theory by the tradition of biological autonomy. It considers organization between parts and processes as the invariant that captures the distinctive character of biological systems. Pioneering conceptual and methodological intuitions in this respect can be found in Rashevsky's work (1954), specifically in his emphasis on relations as what allows identifying a system as a living one, and on the thesis that there is a core set of relations that is common to all living systems.

The autonomy approach has been developed with the aim of identifying and understanding the nature and form of these relationships. This theoretical framework was built upon pioneering work carried out by Jean Piaget (1967), Robert Rosen (1972), Humberto Maturana and Francisco Varela (Varela et al., 1974), Howard Pattee (1972), Tibor Gánti (1975), among others. Recently it was further developed by Stuart Kauffman (2000) and by Alvaro Moreno and collaborators (Moreno & Mossio, 2015). This approach characterizes a biological organism as an autonomous system capable of producing its own components and maintaining itself far from equilibrium with its environment. To explain this capacity, this tradition appeals to the internal organization of the organism, which is maintained in spite of the continuous transformations that an organism undergoes at the level of components.

The core feature of this approach is the focus on the organization of the system. Organization refers to the way production and transformation processes are connected so that they are able to synthesize the very components that realize them, by using energy and matter from the environment. In this view, the fundamental feature of the organization of (biological) self-maintaining systems is its circular topology as a network of processes of production of components that in turn realize and maintain the network itself. This distinctive type of generative circularity that characterizes biological systems is known as 'organizational closure' (Piaget, 1967).

This tradition answers at a different level Schrödinger's question on life and on the invariant order that characterizes it: the abstract level of relationships between parts and processes instead of the level of the intrinsic properties of parts. As argued by Rosen, the idea of a circular invariant organization "looks very much like an aperiodic solid, and indeed it possesses many of the properties Schrödinger ascribed to that concept. The novel thing is that it is not a "real" solid. It is, rather, a pattern of causal organization" (Rosen, 2000: 23).

These ideas have important epistemological implications. The first concerns what level of description is considered as the more pertinent to understand a biological system: material parts, and therefore a bottom-up approach, or relationships, and therefore a top-down approach. While not excluding the first, the autonomy approach focuses mainly on the second, and characterizes the constituents of a biological system in terms of their dependence on and contribution to the system that harbors them: as functional components rather than material parts (Bich, 2012). I will come back to this point in the following sections.

Another more general epistemological implication concerns the descriptive approach developed in the autonomy framework, which is based on organization. The organization of a system is defined as the topology of relations which allows scientific observers to identify a system as a unity belonging to a certain class, that is, the class of living systems (Maturana, 1988). Such a definition entails the impossibility of giving distinctions for granted and of considering scientific knowledge as independent from the activity of observation and categorization performed by an observer. The underlying idea, which is of particular interest when focusing on theoretical and formal modelling, is that an object studied by science is co-construed: the observer gives it an objectual form through the categories she resorts to, while the world, limiting the range of their applicability, defines the area in which nature can be handled in terms of those objects categorized by the observer (Damiano, 2012).

It is important to emphasize a couple of points. In the first place, this epistemological thesis does not mean that categorizations are arbitrary. It is the opposite; they are constrained by interactions with the world and they need to be theoretically coherent. The second is that when focusing on living organisms as self-producing and self-maintaining systems, whose existence and activity coincide, one makes a special type of distinction. An observer identifies a living system as autonomous in the same domain where the system specifies it through its internal operations.

The theoretical and epistemological dimensions of biology are therefore distinctive, or “special”, if life is addressed from the point of view of autonomy. In the next sections I discuss the implications of these ideas.

## 3 Biological Emergence and the Autonomy of Biology

Looking at living systems from the perspective provided by the autonomy framework has deep implications for an understanding of biology and its relationships with other domains of scientific investigation. The first derives from the identification of a causal regime that is distinctively biological: that of organizational closure, according to which biological systems are capable of producing their own components and maintain themselves, unlike other classes of natural and artificial systems (Moreno & Mossio, 2015). The focus of the autonomy framework is primary on the self-maintaining organization of the system and on the consequent idea that the existence and activities of parts depend on such organization. This approach centers on organization and activities of parts (i.e. interactions between parts, operations of parts on processes). The starting point is not components themselves. More specifically, organization refers to the way production and transformation processes are connected so that they are able to synthesize the components that realize them by using energy and matter from the environment. Components are characterized functionally in terms of their activities within a given organization. Studying the role of parts within the system, therefore, needs to take into account the type of organization that harbors them.

This approach raises the problem of how to describe such a highly integrated system and how to decompose it into its parts in order to understand its internal functioning. Surely, a living organism is not an ‘aggregative system’ or ‘component system’ (see Wimsatt, 1986; Bechtel & Richardson, 1993/2010). The parts that contribute to biological phenomena of interest cannot be easily localized, and their activities cannot be considered as fully determined (once their triggering conditions are met) by their intrinsic properties.

Instead, living systems can be considered as a type of ‘integrative system’ or ‘semi-decomposable’ system (Bechtel & Richardson, 1993/2010 ): a class of systems in which the organization contributes to determine the activities of the parts, and the actual results of such activities depend on how their functioning is orchestrated. More precisely, living autonomous systems might be considered as *highly integrated systems*, because not only the activities are ordered in such a way as to achieve specific results, but: (1) the parts depend on one another and on the system for their existence and (2) their activities are not regular and depend on the operations of other regulatory components within the system, which modulate the activities of parts on the basis of the state of the system and the environment (Bich et al., 2016). Decomposing such systems and identifying the parts that are relevant to understand a specific phenomenon or mechanism is a complex endeavor, as showed for example by the history of the discovery of cell mechanisms, metabolic cycles, etc. (see Bechtel, 2006).

For these reasons, in principle an approach based on autonomy privileges decomposing strategies that proceed top-down from the system to the components that contribute to its activities (Rosen, 1991). One way to do so is functionally: to identify one or more activities that are necessary for the organisms to maintain itself and establish which type of operations are necessary to carry it out, and then to identify and characterize in terms of these operations the parts that realize them. It is what Rosen calls ‘analysis’ (Rosen, 1991). The alternative would be to identify the anatomic components of the system, to study them and to use them as the starting point to conceptually reassemble the system. It is what Rosen calls ‘synthesis’ (Rosen, 1991).

The relationships between top-down and bottom-up descriptions—i.e. trying to establish a correspondence between functional parts (identified and characterized top down), and structural ones (anatomic or material parts characterized through a bottom-up approach)—is highly problematic (Bich, 2012). Privileging either approach may result in shortcomings. An exclusive focus on top-down approaches might result in a functionalism characterized by an excessive degree of abstraction and lack of relevant concrete details (Levy & Bechtel, 2013). The risk is to overlook the importance of materiality and of physical aspects to understand how a living system is actually realized. The other way around, a bottom-up approach might lose sight of the causal regime that characterizes the system, with problems of selecting which properties of components are pertinent or not, what components are relevant to describe how a phenomenon is realized, how they behave in different ways depending on the state of the system (Bich et al., 2016, 2020), and when they are not working properly (Saborido & Moreno, 2015; Bechtel, 2018). While irreducible, at least in practice, these two approaches need to proceed hand in hand.

Moreover, it is important to point out that some of the challenges faced in describing biological autonomous systems is that some of the elements needed to define the dynamics of the system are determined, at least in part, from within. By interacting with the environment and establishing their own internal environment, living systems contribute to determine some of the *boundary conditions* that allow them (and their internal processes) to exist; they also determine and modulate some of the *parameters* of their internal dynamics, for example by activating or inhibiting the activity of enzymes and regulatory proteins, and finally, the *rules of interaction* between parts depend on what components are produced and how their operations are modulated by regulatory mechanisms (Kampis, 1991; Bich & Bocchi, 2012; Longo et al., 2012; Koutroufinis, 2017). One may also argue that the organization of the system has a role in determining or constraining the behavior of the parts (El-Hani & Queiroz, 2005; Mossio et al., 2013).

The idea of a distinctive causal regime of self-maintenance and the epistemic implications it brings to surface with regards to the study of biological systems, have often led to associating biological autonomy with emergence (Rosen, 1991; Varela, 1997; Kauffman, 2000; Bich, 2006; Mossio et al., 2013). The causal irreducibility of the regime of organizational closure is paired with an epistemological irreducibility: a number of limitations regarding the possibility of understanding, modelling and formalizing these types of systems. Different types of descriptions (such as in term of material and functional parts) coexist, and so do irreducible notions (such as sequence and function).

In general, there is no preferential or more pertinent heuristics in general. What types of heuristics one needs to adopt may depend on the specific phenomenon under investigation. For example, let us think about the strategies employed in the discovery of physiological processes such as fermentation and oxidative phosphorylation, which sought the opposition between reductionist and anti-reductionist approaches, with competing agendas and heuristics (Bechtel & Richardson, 1992). In the late nineteenth century, while reductionists associated fermentation with independent chemical reactions, anti-reductionists such as Schwann and Pasteur argued that fermentation required taking into consideration the circumstances found in living cells instead of looking *only* to parts. The discovery of the mechanisms of fermentation happened in several steps in a period spanning from the last decades of the nineteenth and to the first ones of the twentieth centuries. It did not result in the reduction of this phenomenon to a chain of reactions, although reactions had to be identified. Looking also at the types of connections between the reactions involved resulted in the discovery of an organized biochemical system characterized by several causal loops: “a highly integrated, interlocking *system* of reactions” (Bechtel & Richardson, 1992: 273).

Similarly, the discovery of oxidative phosphorylation also showed the difficulties of identifying the pertinent levels for explaining the phenomenon under investigation. It required considering not only individual reactions and their dynamical organization, but also including structural aspects at a different scale than that of chemical reactions, such as the macroscopic structure of the mitochondrion and in particular of its systems of endomembranes, which were studied through electronic microscopy. Moreover, as shown by Bechtel and Richardson (1992), techniques that were relevant for one level of descriptions often concealed or even destroyed crucial aspects of the phenomena investigated, which were only available, instead, at other levels of description.

Many of these considerations, apart from those derived from the notion of organizational closure, are not exclusive of biological systems. Physics provides examples and formal models of natural phenomena that can be considered as emergent and that cannot be predicted or deduced from a description of their constituents or from an initial state. These phenomena, therefore, are described by employing models that are irreducible to one another (Pessa, 1998). From physics itself comes a questioning of the very idea of a fundamental level of description and of fundamental objects (Pessa, 2011; Bitbol, 2007).

However, focus here, from a systemic perspective, on a few elements which, among others, are distinctively biological and ground some degree of autonomy for this discipline. They are useful in order to discuss then the epistemic status of biology. Among the differences between the domains of physics and biology, Longo and Montévil (2013), have discussed what are the features of the objects characteristic of either domain and how to describe their behaviors. Physical objects are generic because different objects of the same category have the same intrinsic features and behave in the same way. Their trajectories in phase space are instead specific and defined by the relative equations. For biological objects, the opposite is true: they are specific while their behavior is generic as they follow a possible evolutionary trajectory in the phase space.

What about some core biological concepts such as teleology, function, integration, regulation, control, neither of which has a counterpart in physics nor has been reduced to physical concepts? These concepts make biological explanation theoretically independent from the physical and chemical ones. They enable explanations that are directly grounded in the specificity of biological phenomenology rather than derived from lower-level explanations. These concepts may constitute heuristic tools that are useful to address biological phenomena in practice, but different attempts have also been made to naturalize them and make them well-grounded theoretical notions.

One interesting case is that of processes oriented towards a final state. In physics one can find, among others, the Geodesic Principle, Le Chatelier's principle and the Second Law of Thermodynamic, which describe how the trajectories of certain systems tend to proceed towards a final state such as for example, thermodynamic equilibrium. Biological systems exhibit a similar yet qualitatively distinct feature: they actively pursue certain states, which can be considered the goals of the system. What for physical systems are end states, for biological ones become goals, aims, purposes. All these notions, which belong to the category of teleology, are not just ways of speaking or heuristic tools, useful to describe the behavior of a system, but can be provided a naturalized grounding in the autonomous organization of living systems (Mossio & Bich, 2017, see also Schlosser, 1998; Delancey, 2006). A living system is characterized by the distinctive capability to produce, transform and repair its components which realize and maintain the system through its interactions with the environment. Its own activity and those of its parts are, in a fundamental sense, oriented toward an end. The goal of the system is to maintain itself. It is true that there are other systems, among artifacts, which are considered as goal-oriented in their behavior. An example is a thermostat, which controls the temperature of a room to maintain it within a certain interval of temperature. Yet there is a fundamental difference between this type of goal-oriented behavior and the teleological one of biological systems. It amounts to the difference between *following* or *having a goal* (Jonas, 1953). Artifacts *follow a goal*. The goal of artifacts is determined *extrinsically*, by the designer or the user. Following this goal does not contribute to the existence of the artifact. Biological systems "act on their own behalf" (Kauffman, 2000). They *have* an *intrinsic goal*, which is their own existence.

Another fundamental concept for biology is that of function, which also has a teleological dimension. Functional explanations are widespread in biology, and parts and traits of living systems are often characterized in terms of what they do. There are general accounts of functions, such as the dispositional one, which are generic and can be applied to almost any class of system (Cummins, 1975). The dispositional approach identifies the function of a part with its causal role in a larger system. Yet the generality of this concept sacrifices other aspects which are important in developing a scientific explanation. Ascribing functions to a part in terms of causal role may be arbitrary and it may not provide a normative basis for distinguishing which among many causal effects to count as the function of a component.

Nevertheless, there are principled way to ascribe functions that are specifically biological (i.e. capture the distinctive and irreducible character of biological functions) and, unlike the dispositional account, justify claims such as that the function of the heart is to pump blood (and not, for example, to make noise). The most widespread account is based on evolutionary considerations, and characterizes a function as a selected effect of a trait of an organism which contributed to the survival of the ancestors of that organism (Millikan, 1989; Neander, 1991). The autonomy framework, instead, characterizes functions in terms of contributions to the maintenance of the organism. A function is understood as a contribution of a trait to the maintenance of an autonomous organization (e.g., a living cell) that, in turn, contributes to producing and maintaining the trait itself (Collier, 2000; McLaughlin, 2001; Christensen & Bickhard, 2002; Mossio et al., 2009). The way functional ascriptions are justified and employed in biology does not have counterparts in physics and chemistry.

## 4 Is Biology Marginal? Insights for an Epistemology of Complex Systems

The previous sections have discussed the theoretical account of living systems based on the notions of biological autonomy and organization. They showed how, if one adopts this perspective, biological phenomena can be considered emergent from the causal and epistemological points of view. In this scenario, the biological domain can be considered as exhibiting distinctive phenomena and requiring concepts that have no counterpart in other scientific domains that focus on lower levels of organization.

However, biological phenomena are rarer than physical and chemical ones, and biology as a discipline concerns distinctive phenomena, exhibits a certain degree of autonomy (although not self-sufficiency) with respect to other sciences, and employs its own concepts. Does this mean that biology is marginal, as argued by Monod (1970)? And is there a wider lesson to be learned from biology? These questions can be addressed in multiple ways by either focusing on theoretical aspects or epistemic ones. I argue that, from both points of view, the answer is that biology is not marginal and there are lessons to be learned from it.

One way to approach these issues is theoretical. If the tools of sciences such as physics and chemistry, although useful and unavoidable, are unable to provide an understanding of what living systems are and how they function, one needs to expand science. This answer is in line with claims such as the one made within philosophy by Hans Jonas. When discussing life, he argued that "if life is not within the competence of an alleged cosmic principle, though it is in every sense within the cosmos, then that principle is inadequate for the cosmos as well" (Jonas, 1966: 65). The research projects carried out by Schrödinger (1944) and Kauffman (2000)—aimed respectively at developing a new physics, or new laws, to make sense of living systems as natural phenomena—constitute attempts to respond to these questions within science. These attempts aim to extend a cohesive set of theoretical tools. Another possibility, more in line with the autonomy framework, is to complement the tools provided by physics and chemistry with new tools, such as organizational closure, specifically developed for addressing living systems. In the first case, biology would not be marginal because it would be part of an extended, more general, physics; in the second case because it would be source of new theoretical tools and principles applied in combination with those of physics and chemistry: a more general science. In both cases, biology would be a source of new lessons for science.

Acknowledging the distinctive character of biological systems implies on the one hand the idea that the biological domain should not be considered as a particular case of other domains considered as more fundamental, such as physics and chemistry. Biological systems can be investigated in their specificity only by building new types of theoretical and descriptive models. On the other hand, it makes it necessary to consider the natural world as characterized by a range of phenomena much wider and richer than what can be addressed through the tools of one discipline or approach alone, be it biology or physics. This, according to Rosen, is one of the meanings of Schrödinger's insight on a new physics, which becomes the foundation for a theoretical research program for biology and for complexity sciences in general. According to Rosen, Monod's thesis on the marginality of biology rests on the idea that organisms are "just specializations of what is already on the shelf provided by old physics, and that to claim otherwise is mere vitalism" (Rosen, 2000: 26). Organisms are indeed rare if compared to other material systems. Yet Rosen argues that Monod's argument builds upon an artifact of sampling: a confusion between 'rare' and 'special' (in the sense of marginal).

In sum, from the theoretical point of view, organisms are more general, insofar as they exhibit properties and phenomena that require the development of new conceptual categories, capable to capture also those aspects, such as closure, teleology, functionality, regulation, control, that escape other conceptual frameworks. There is something qualitatively different in biological systems, invisible to other sciences, and that requires a conceptual rethinking and new categorizations to be employed together with those derived from other sciences such as physics or chemistry. In this sense the study of living systems and of their distinctive character carries a lesson on nature and science in general.

The other way to address the questions is to focus on epistemic aspects, and to consider biological systems as paradigmatic cases of complex systems. This epistemological thesis is specifically connected to the problem of the relationships between scientific disciplines and descriptive strategies. It supports a non-reductionistic approach oriented towards establishing of communicative circuits between disciplines and between different heuristics within a discipline: biology as a model for an epistemology of complex systems characterized by different irreducible approaches and descriptive tools which coexist and interact; a domain characterized by multidirectional transfers of models, questions and theoretical structures.

Let us consider some issues deriving from the difficulty of establishing connections between different types of observations, models, observables, etc. As discussed in Sect. 2, the relationships between directions of observation is one of these cases. In the autonomy approach the observables that are built bottom-up from the observation of intrinsic properties of the parts and those built top-down in terms of functional properties (identified with regards to the contribution of components to the system) do not necessarily coincide (Rosen, 1991; Bich, 2012). Material and functional components may not be one and the same thing. An example is the case of enzymes. A bottom-up analysis in terms of sequence of amino acids may not covey the same information as a structural analysis of the configuration and functionality of the folded molecule. For the same sequence there might be several configurations, depending on the boundary conditions present during folding, the activity of chaperons, and of several regulatory interactions such as phosphorylation and allosteric control. A mixed approach is often fruitful to predict possible regulatory sites, and how interactions at these sites changes the probability of having a given configuration (and functional capability) of the molecule.

Moreover, differences in the types and scales of observation may provide different pictures of the biological phenomenon under investigation. Let us think again of the discovery of phenomena such as fermentation and oxidative phosphorylation (Bechtel & Richardson, 1992). Whereas a study of individual chemical reactions, or sequences of reactions, was an important aspect, it proved to be insufficient to provide an understanding of these two phenomena. Some gestalt switches were needed. In the case of fermentation, a different type of perspective was needed, focused on the topology of the relationships between the reactions, i.e. their organization, which led to the discovery of chemical cycles. The case of oxidative phosphorylation showed the importance of taking into account different irreducible scales by complementing the investigation of chemical reactions with that of the role of macroscopic structures such as membranes, which constrain these reactions and enable different types of processes.

Similar considerations can be made with respect to general strategies employed to describe and model living autonomous systems. As argued by Moreno and Suárez (2020), two irreducible strategies provide information on different aspects of the system. One is network modelling: a holistic tool used to study and predict global dynamical properties of large sets of interacting entities. It has often been employed to investigate the dynamical properties (e.g. stability and robustness) of abstract theoretical models of the organization of biological autonomous systems (Piedrafita et al., 2010). The other strategy is the new mechanistic one, which aims to provide a causal explanation of how the individual parts of the systems, or the parts of one or several subsystems, functionally operate and interact within autonomous systems to realize specific phenomena. This strategy has been recently applied to model and analyze phenomena such as mammary organogenesis (Montévil et al., 2016) and glycaemia regulation (Bich et al., 2020) from an autonomy perspective. Network and mechanistic strategies provide different information on the system. Although irreducible to each other, they can be combined. Network modeling, for example can be used to support mechanistic descriptions. Identifying the most connected nodes of a network may provide insight into what may be the relevant functional components responsible for producing the phenomenon under investigation (Bechtel, 2015).

These examples show both the importance of considering how the system is organized in different layers—which instantiate distinct and complementary descriptive domains—and to take into account the role of the observer who needs to adopt different modalities of description in order to account for them. The common aspect to these examples is that sets of models derived from different observational operations or descriptive strategies provide different, though complementary, information about the system under study. They show the failure or the inadequacy of a single descriptive modality and the consequent necessity to include new ones. Some modalities might be more relevant or pertinent than others depending on the phenomenon to study and the aims of the scientist, but there seems to be no privileged one so that best result are obtained when more strategies are used in combination.

For these reasons it can be claimed that biological systems are emergent from an epistemological standpoint. Emergence in this sense depends on the relationship between different models that are needed in order to describe the system and depends on the experiences performed by an observer who interacts with it. In this framework it can be expressed as the lack of a direct relationship between different descriptions made in distinct domains or different types of descriptions of the same phenomenon (Bich, 2012).

These conclusions, drawn from the discussion of a systemic account of biological systems, have a more general relevance. They are in line with Rosen's epistemological account of complexity, according to which "To say that a system is complex [...] is to say that we can describe the same system in a variety of distinct ways [...]. Complexity then ceases to be an intrinsic property of a system, but it is rather a function of the number of separate descriptions required [ . . . ]. Therefore, a system is simple to the extent that a single description suffices to account for our interaction with the system; it is complex to the extent that it fails to be true." (Rosen, 1978: 112). This notion is focused on the relationship between classes of observables that converge in different models. Complexity can be defined as the insufficiency of a single model, and of the set of observables related to it, to describe a system. Consequently, one model needs to be replaced or complemented by other ones, because the system exhibits to the observer new characteristics which were not present before or at a different level of description, and which are thus invisible to the chosen observables that constitute the starting description.

This perspective is also in line with a heuristic of complex systems such as the one proposed by Minati, Penna and Pessa (Minati et al., 1998; Minati & Pessa, 2006), based on the dynamical usage of models: the interaction between different, and often complementary, models which work in the traditional way just inside their limited domain of validity. This means not just that one needs to choose an individual model as the more appropriate in order to address a specific issue, but also that to investigate complex systems one may need to employ more than one descriptive modality at the same time and multiple interacting models.

On this basis, one may argue that biology is general insofar as it constitutes a paradigmatic example of complexity science, both in terms of how it defines the theoretical object of study and of the epistemology and heuristics employed. Addressing biological phenomena from an autonomy perspective brings to the attention in a wider context the limitations of approaches based on simple systems and the virtues of adopting ones based on complexity. Focusing on organization, function, teleology and other biological notions, for example, shows how scientific investigation needs to combine analytic and synthetic strategies. Understanding what makes a system a living organism, which exhibits distinctive features with respect to physical and chemical systems, has relevant consequences not just for those "rare" phenomena pertaining to biology but for scientific explanation in general, more so for the study of complex systems. It provides theoretical and epistemological grounds to advocate a pluralist perspective combining different points of view and descriptive and explanation strategies.

Therefore, addressing question "what is life?", that is, the problem of defining and characterizing living systems, does not consist only in responding to the needs of one discipline. It has wider consequences, or lessons, and introduces more general questions that cut across scientific domains. Focusing on biological systems and their specificities shows how complex phenomena may escape individual strategies, how general fundamental issues related to complex systems such as the notion of system as an integrated organized entity, the relationship between wholes and parts need to be addressed in more than one way and direction. The challenge, each time, is to understand how to combine these different tools and strategies rather than extend a given one or choosing one among many.

## 5 Conclusions: A Practice of Complex Systems

The passage from a conceptual discussion of complex systems to their study and modelling in practice, from theory to art one might say, is not direct and it may be quite difficult. This is especially relevant for philosophers, whose goals are often generality and abstraction, but who constantly face the risk of the excess of idealization, of striving for clear-cut concepts and distinctions that as a result may be too detached from scientific work. In the study of complex systems, where the activity of scientists and its limitations play a crucial role, this might create a gap between epistemological thought and actual modelling, and even lead to naïve conclusions about the relation between theoretical science, modelling, and the natural world. In particular, the study of theories and models of complex systems requires tools and heuristics to identify and analyze their limitations and to discuss how such limitations can be faced by employing multiple strategies.

In this context, virtuous examples are fundamental to develop epistemological thinking. I consider myself honored and lucky for having had the opportunity to know Eliano Pessa as teacher, supervisor and then colleague. He guided me through my first steps into complex systems thinking from the point of view of science. Not only he introduced me to the notion of emergence and to the theoretical work of Rashevsky, Rosen and of Maturana and Varela, but with his generosity and honesty he gave a virtuous example of the art of studying and modelling complex systems, with all the difficulties and stimulating challenges that characterize this practice, and he transmitted his enthusiasm to students and colleagues. In particular, I remember his capability of making explicit the idealizations underlying models and discussing their implications, of explaining with incredible clarity the ingenious models he developed while at the same time always showing their limitations with irony, precision and detail. With his example, Eliano Pessa demonstrated that striving for honesty is a fundamental epistemic value in scientific research and showed the importance of giving substance to epistemological and theoretical thinking.

**Acknowledgments** The author acknowledges funding from the Ministerio de Ciencia, Innovación y Universidades, Spain ('Ramon y Cajal' Programme RYC-2016-19,798 and research project PID2019-104576GB-I00) and the Basque Government (Project: IT1228–19).

## References

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge University Press.

Bechtel, W. (2015). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 53*, 84–93.

Bechtel, W. (2018). The importance of constraints and control in biological mechanisms: Insights from cancer research. *Philosophy of Science, 85*, 573–593.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.

Bechtel, W., & Richardson, R. C. (1992). Emergent phenomena and complex systems. In A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 257–288). de Gruyter.

Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press (1993 edition published by Princeton University Press).

Bich, L. (2006). Autopoiesis and emergence. In G. Minati, E. Pessa, & M. Abram (Eds.), *Systemics of emergence. Research and development* (pp. 281–292). Springer.

Bich, L. (2012). Complex emergence and the living organization: An epistemological framework for biology. *Synthese, 185*(2), 215–232.

Bich, L., & Bocchi, G. (2012). Emergent processes as generation of discontinuities. In G. Minati, E. Pessa, & M. Abram (Eds.), *Methods, models, simulations and approaches towards a general theory of change* (pp. 135–146). World Scientific.

Bich, L., & Damiano, L. (2008). Order in the nothing: Autopoiesis and the organizational characterization of the living. *Electronic Journal of Theoretical Physics, 4*(16), 343–373.

Bich, L., & Green, S. (2018). Is defining life pointless? Operational definitions at the frontiers of biology. *Synthese, 195*(9), 3919–3946.

Bich, L., Mossio, M., Ruiz-Mirazo, K., & Moreno, A. (2016). Biological regulation: Controlling the system from within. *Biology and Philosophy, 31*(2), 237–265.

Bich, L., Mossio, M., & Soto, A. (2020). Glycemia regulation: From feedback loops to organizational closure. *Frontiers in Physiology, 11*, 69.

Bitbol, M. (2007). Ontology, matter and emergence. *Phenomenology and the Cognitive Science, 6*, 293–307.

Christensen, W., & Bickhard, M. (2002). The process dynamics of normative function. *The Monist, 85*(1), 3–28.

Collier, J. (2000). Autonomy and process closure as the basis for functionality. *Annals of the New York Academy of Science, 901*, 280–290.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy 72*, 741–765.

Damiano, L. (2012). Co-emergences in life and science: A double proposal for biological emergentism. *Synthese, 185*, 273–294.

Delancey, C. S. (2006). Ontology and Teleofunctions: A Defense and Revision of the Systematic Account of Teleological Explanation. *Synthese, 150*(1), 69–98.

El-Hani, C. N., & Queiroz, J. (2005). Downward determination. *Abstracta, 1*(2), 162–192.

Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: The chemotons. *Biosystems, 7*, 15–21.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

Jonas, H. (1953). A critique of cybernetics. *Social Research, 20*, 172–192.

Jonas, H. (1966). *The phenomenon of life. Towards a philosophical biology*. Harper and Row.

Kampis, G. (1991). *Self-modifying systems in biology and cognitive science*. Pergamon Press.

Kauffman, S. A. (2000). *Investigations*. Oxford University Press.

Koutroufinis, S. A. (2017). Organism, Machine, Process. Towards a Process Ontology for Organismic Dynamics. *Organisms. Journal of Biological Sciences, 1*(1), 23–44.

Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science, 80*, 241–261.

Longo, G., & Montévil, M. (2013). Extended criticality, phase spaces and enablement in biology. *Chaos, Solitons & Fractals, 55*, 64–79.

Longo, G., Montévil, M., & Kauffman, S. (2012). No entailing laws, but enablement in the evolution of the biosphere. In *Proceedings of the Fourteenth International Conference on Genetic and Evolutionary Computation Conference Companion—GECCO Companion '12* (p. 1379). https://doi.org/10.1145/2330784.2330946.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Maturana, H. (1988). Reality: The search for objectivity or the quest for a compelling argument. *The Irish Journal of Psychology, 9*(1), 25–85.

McLaughlin, P. (2001). *What functions explain. Functional explanation and self-reproducing systems*. Cambridge University Press.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302.

Minati, G., Penna, M. P., & Pessa, E. (1998). Thermodynamical and logical openness in general systems. *Systems Research and Behavioral Science, 15*(2), 131–145.

Minati, G., & Pessa, E. (2006). *Collective beings*. Springer.

Monod, J. (1970). *Les hasard et la necessité*. Seuil.

Montévil, M., Speroni, L., Sonnenschein, C., & Soto, A. M. (2016). Modeling mammary organogenesis from biological first principles: Cells and their physical constraints. *Progress in Biophysics and Molecular Biology, 122*, 1–12.

Moreno, A., & Mossio, M. (2015). *Biological autonomy: A philosophical and theoretical enquiry*. Springer.

Moreno, A., & Suárez, J. (2020). Plurality of explanatory strategies in biology: Mechanisms and networks. In W. J. Gonzalez (Ed.), *Methodological prospects for scientific research* (pp. 141–165). Springer.

Mossio, M., & Bich, L. (2017). What makes biological organisation teleological? *Synthese, 194*(4), 1089–1114.

Mossio, M., Bich, L., & Moreno, A. (2013). Emergence, closure and inter-level causation in biological systems. *Erkenntnis, 78*(2), 153–178.

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *British Journal of Philosophy of Science, 60*(4), 813–841.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defence. *Philosophy of Science, 58*, 168–184.

Pattee, H. H. (1972). The nature of hierarchical controls in living matter. In R. Rosen (Ed.), *Foundations of mathematical biology. Volume I: Subcellular systems* (pp. 1–22). Academic Press.

Pessa, E. (1998). Emergence, self-organization, and quantum theory. In G. Minati (Ed.), *First Italian Conference on Systemics* (pp. 59–80). Apogeo.

Pessa, E. (2011). The concept of particle in quantum field theory. In I. Licata & A. Sakaji (Eds.), *Vision of oneness* (pp. 13–40). Aracne.

Piaget, J. (1967). *Biologie et Connaissance*. Gallimard.

Piedrafita, G., Montero, F., Moran, F., Cardenas, M. L., & Cornish-Bowden, A. (2010). A simple self-maintaining metabolic system: Robustness, autocatalysis, bistability. *PLoS Computational Biology, 6*(8), e1000872.

Popa, R. (2004). *Between necessity and probability: Searching for the definition and origin of life*. Springer.

Rashevsky, N. (1954). Topology and life: In search of general mathematical principles in biology and sociology. *Bulletin of Mathematical Biophysics, 13*, 317–348.

Rosen, R. (1972). Some relational cell models: The metabolism-repair systems. In R. Rosen (Ed.), *Foundations of mathematical biology. Volume II: Cellular systems* (pp. 217–253). Academic.

Rosen, R. (1978). *Fundamentals of measurement and representation of natural systems*. North Holland.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.

Rosen, R. (2000). *Essays on life itself*. Columbia University Press.

Saborido, C., & Moreno, A. (2015). Biological pathology from an organizational perspective. *Theoretical Medicine and Bioethics, 36*, 83–95.

Schlosser, G. (1998). Self-re-production and functionality: A systems theoretical approach to teleological explanation. *Synthese, 116*, 303–354.

Schrödinger, E. (1944). *What's life? The physical aspect of the living cell*. Cambridge University Press.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition, 34*, 72–87.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems, 5*(4), 187–196.

Wimsatt, W. C. (1986). Forms of aggregativity. In A Donagan, , A. N. Perovich, M. V. Wedin (Eds.), Human nature and natural knowledge (pp. 259–291). Reidel.

# Information from Structure: How Networks Face Biological Complexity

**Alessandro Giuliani**

**Abstract** The multi-level organization of nature is self-evident: proteins do interact among them to give rise to an organized metabolism while in the same time each protein (a single node of such interaction network) is itself a network of interacting amino-acid residues allowing coordinated motion of the macromolecule and systemic effect as allosteric behaviour.

Similar situations hold for ecology, anatomy, organ physiology. The most diffuse approach to such situation is to give for granted that causally relevant events pertain to the most fundamental level (the molecular one) in the form of regularities (or perturbations in the case of pathological situations), that 'climb up' the hierarchy reaching the ultimate layer of macroscopic behaviour.

Such causative model, is not the only one: we observe top-down, bottom-up as well as middle-out perturbation/control trajectories.

The recent complex network studies allow to go further the pure qualitative observation of the existence of both non-linear and non-bottom-up processes and to uncover the deep nature of multi-level organization. Here, taking as paradigm protein science, we will give an account of how the information travelling across a network can create meaning so offering a more realistic frame of causation in complex systems.

**Keywords** Allosteric effect · Biodynamic interfaces · Causative models · Complex networks · Complexity · Multi-level organization · Proteins

## 1 Introduction

The network formalism is probably the most natural way to represent biological systems. Even if in the last decades the analysis of complex networks became a

---

A. Giuliani (✉)
Environment and Health Department, Istituto Superiore di Sanità, Rome, Italy
e-mail: alessandro.giuliani@iss.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_5

very widespread paradigm to face problems going from macromolecular structures (Di Paola et al., 2013) to genetic regulation circuits (Demongeot et al., 2003), neuroscience (Sporns, 2018) and ecological systems (Mendonça et al., 2018) this is not a new idea. In 1948 Warren Weaver (1948) one of the fathers of mathematical information theory sketched a very intriguing synthetic tripartite description of science into problems of 'organized simplicity', 'disorganized complexity' and 'organized complexity' with biology located in the last class.

The first class (simplicity) refers to the case of very few elements interacting among them with largely invariant relations, being its paradigm classical mechanics. Class 1 problems allow for an extreme abstraction (e.g. a planet can be thought as a dimensionless 'material point'). The possibility to take into consideration only very few basic features like mass and distance makes this approach largely object independent (this is the basic reason of the use of examples of the same physical law based upon cars, cannonballs or skiers).

Problems of Disorganized Complexity (class 2) have as paradigm classical thermodynamics and reach a great generalization power by means of a very different style of reasoning with respect to Class 1. In Class 2 problems, the predictive power stems from giving up the analysis of system fundamental aggregation level (e.g. the single molecules) preferring a statistical knowledge corresponding to gross averages (like pressure, volume, temperature are) on a transfinite number of atomic elements. Both the approaches must fulfil very stringent constraints. Class 1 approach asks for few involved elements interacting in a stable way, class 2 style needs a very large number of identical particles with only negligible (or very stable and invariant) interactions among them. Biological systems, only in very few cases do satisfy these constraints, so we step into Weaver's third class (Organized Complexity). Organized Complexity arises whenever many (even if not so many as in class 2) non-identical elements each other interact by means of links endowed with time-varying correlation strength.

The interaction of 'non identical elements' with 'varying correlation strengths' corresponds to a network of links (edges, correlations) with variable strength, connecting different nodes that in turn are 'non identical' being themselves networks with variable wiring structure.

Weaver (1948) commented that while science was at home (relying on the usual repertoire of laws and boundary conditions deciding for their application) in both Class 1 and Class 2 phenomena, the overwhelming importance of contextual information with respect to lawful invariant behaviour, of Class 3 systems, makes the situation much more uncomfortable. After more than 70 years from Weaver's paper, we made some steps ahead in Organized Complexity studies and the present work deals with these advancements.

The paper is organized as follows: in the first chapter (biodynamic interfaces) we will discuss the basic principles of the interaction between complex systems, with an emphasis on the need of an intermediate layer shared by the two interacting systems with a partially independent nature with respect to the two interactors. In the second chapter (the middle way) we will introduce the concept of mesoscopic or 'middle-out' organization demonstrating why the 'network representation' allows for a natural, hypothesis free formalization of the meso-scale. The third chapter will be devoted to the transit of information across a network system and the consequent discrimination from noise of the relevant (signal) perturbations able to 'climb-up' or 'stepping-down' the multilevel organization using allosteric effect in proteins as model system.

## 2 Biodynamic Interfaces

There is no interaction without information exchange and there is no information exchange without an efficient communication channel. This 'channel' is exactly what we call 'Interface'. If Mary calls Peter by means of her smartphone, the establishing of a contact strictly depends on the existence of an electromagnetic field endowed with a band of frequencies devoted to cell phone communication. Peter smartphone corresponds to a very specific frequency modulation of the field that is elicited by the digits Mary composes on her phone and sends on the specific band of frequencies, consequently Peter's smartphone rings and the communication begins. We do not enter into the actual content of communication (that only pertains to Mary and Peter), instead we focus on two crucial points of the process:

1. The existence of a medium (the field) that cannot be considered as a discrete entity with a specific location in both space and time but as a 'global feature' covering the space and assuming different values in different locations. The interactors (here the Mary and Peter phones) are causally linked in both directions only because they share the same field. From basic physics we know that a point charge embedded into an electromagnetic field both 'senses' (i.e. is influenced by the field) and modifies (i.e. influences) the field. This is exactly what happens in human-environment interaction in which environment influences physiology (e.g. toxic effects, sensory information..) and is in turn influenced by humans. Both human beings and environment are complex systems and, for their interaction, they need a shared interface (Arora et al., 2020).
2. The interface (field) oscillates with a specific frequency, this implies it has both a 'spatial' and a 'temporal' structure, it is a dynamic interface. The frequency of oscillation is not independent from the spatial features of the interface, more in general, any network system (even a field can be imagined as a grid with some focal points, the 'cells' in the case of mobile phones) has characteristic oscillation modes originating from its wiring structure. We will go back on this point when dealing with protein structures 'resonating' with specific modes that are the carriers of across levels information.

Both these issues are at work in multi-level organization and, more in general, in biological regulation.

## 3 The Middle Way

The most common style of explanation follows an IF-THEN style in which what happens at a given level influences (or determines depending on the relative importance of stochasticity embedded in the link). These linear fluxes of implications give for granted the existence of a fundamental 'explanatory layer' located at the most microscopic level that, thanks to a sort of domino effect, ends up into a macroscopic consequence.

This view is in sharp contrast with what we know about complex structured systems, where a multi-layer (and bi-directional) causality is at work. One of the most clear falsifications of the obliged 'bottom-up' character of biological causation, comes from a 1945 paper (Fankhauser, 1945) by the German (but USA based) embryologist Gerhard Fankhauser. He considered cell size in polyploid triton larvae that have a doubled chromosome number with respect to their diploid counterpart. The polyploid individuals have a doubled cell size with respect to the diploid ones, on the contrary, they have exactly the same dimension of organs and ducts. Arora et al. (2020) with respect to the diploid counterpart. This comes from the fact that the polyploid organism uses half the number of cells, though each cell was itself double in size, to build up its organs. This is crucial for life: the optimization of the calibre of a biological structure (the duct) is fine tuned to fit with the flow of biological fluids (a top-down constraint) and cannot be established by more fundamental levels like its constituent cells or the genome. While this is an intuitive tenet for a 'designed' or 'teleological' process (after all, we do not decide the size of our house based solely on the size of the bricks!), the Fankhauser finding was considered as a largely unexpected finding in a natural system. This is why Albert Einstein (a colleague of Fankhauser at Princeton) told he was expecting the double size cells should give rise to double size organs, concluding that the Fankhauser observation pointed to still largely unknown principles. The brilliant Fankhauser experiment was largely overlooked and obscured by the successes of molecular biology in the years to come, but it is a clear example of a top-down causative model in which a 'high-level' constraint 'slaved' the microscopic cellular/genomic level.

It is important to stress that the 'bottom-up only' obsession is not shared by all the biological fields of investigation, Ecologists recognized since many years that the 'most microscopic' level of organization is not necessarily the place where 'the most relevant facts do happen'.

On the contrary, the most fruitful scale of investigation is where 'non-trivial determinism is maximal' (Pascual & Levin, 1999). That is to say, the scale more 'rich' in meaningful correlations between features pertinent to micro and macroscale that directly recalls the above sketched concept of 'Interface'.

Non-trivial determinism can be defined in terms of prediction error as:

$$\text{Prediction } r^2 = 1 - E^2/S^2$$

In the above formula, E is the mean prediction error and S the standard deviation.

In the case of a simple linear regression in which a dependent variable Y must be predicted by an independent variable X, the non-trivial determinism is nothing else than the usual squared Pearson correlation between the two X and Y variables.

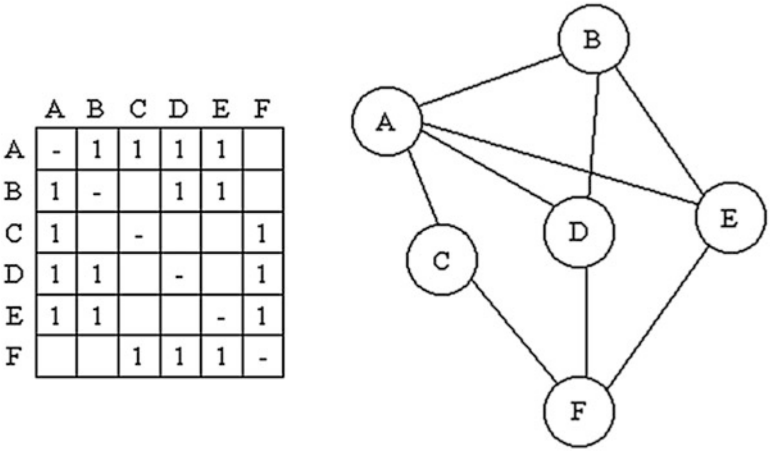
The formula can be extended to any other situation in which we wish to predict a system feature Y, both X and Y do not need to represent single variables but any suitable set of information at any definition scale.

The ‘non-trivial’ attribute of determinism stands for the need of ‘explaining the variance’ of the system at hand (the statistic $r^2$ corresponds to the proportion of variance explained by a model) and not its ‘average’ pattern. The aim is to get rid of the actual behaviour of the system in both space and time and not to describe a ‘frozen’ ideal configuration.

The individuation of ‘mesoscopic principles’ largely independent from the material constitution of the studied system and only dependent from their relational structure was faced by 1998 Nobel prize in Physics Robert Laughlin and colleagues. A paper appeared in year 2000 (Laughlin et al., 2000) entitled ‘The Middle Way’ that aptly individuated in the discovery of universal mesoscopic principles the next frontier of science.

As pointed out by Nicosia et al. (2014): ‘*Networks are the fabric of complex systems*’ and this tells us that network formalism is probably the ideal instrument in the search for such principles. The basic idea of complex network style of reasoning is that shared organization rules (i.e. similar wiring patterns) give rise to similar phenomenology, independently of the nature of the constituting elements. In other words, complex network invariants promise to be the place where to look for universal mesoscopic principles, for the simple fact that they have not different regularities and laws for the different levels, this promises to be the viewpoint that maximizes ‘non-trivial determinism’ (Pascual & Levin, 1999) favoring the emergence of between-level correlations.

In Mickulecki (2001) paper, the author demonstrates the neat separation of the laws governing the internal functioning of the nodes of a network (constitutive laws) from the laws and regularities only dependent from the wiring structure of the system (relational laws). This allows to build an electrical analogue of a mechanical or physiological system only based on conservation principles of both potential and flux across a network analogous to Kirchoff’s laws. The flux does not need to be an electrical current and the same holds for the potential: a system represented by a set of nodes linked by edges with a given topology has similar emerging properties independently of the physical nature of nodes and edges. This opens the way to a ‘network thermodynamics’, whose principles are strictly dependent from wiring architecture while largely independent of the constitutive laws governing the single elements. Still more important, this provokes a shift from the founding of the unitary character of science from the consideration that ‘all the entities are made of the same fundamental building blocks’ to the recognition that ‘all the entities can be represented by a set of relations among their parts’. These relations can be formalized in terms of graph (network) invariants catching different aspects of the wiring structure of the system at hand.



|   | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| A | - | 1 | 1 | 1 | 1 |   |
| B | 1 | - |   | 1 | 1 |   |
| C | 1 |   | - |   |   | 1 |
| D | 1 | 1 |   | - |   | 1 |
| E | 1 | 1 |   |   | - | 1 |
| F |   |   | 1 | 1 | 1 | - |

**Fig. 1** The figure reports the adjacency matrix (left panel) correspondent to the wiring diagram on the right. The presence of a direct link between two nodes corresponds to a unit value of the corresponding element of the matrix on the left. Here all the links are supposed to have the same strength, in other cases we can substitute the unit values with a quantitative estimate of the correlation strength. The represented graph is bi-directional

Complex network invariants catch the essence of multi-level organization for the simple fact their estimation merges different level of definition of the system at hand without the need of any strong hypothesis.

Mathematically speaking, a network corresponds to a graph whose entire information is caught by its adjacency matrix (see Fig. 1): a binary matrix having as rows and columns the nodes and at each i, j position a unit value if the i and j nodes have a direct link between them and 0 otherwise.

Graph invariants are relative to local (single nodes), global (entire network), and mesoscopic (clusters of nodes, optimal paths) levels respectively. The "degree" (how many links are attached to a given node) is a local descriptor, the "average shortest path" (characteristic length) is the average length of minimal paths connecting all the node pairs, and can be considered as a mesoscopic feature, while the general connectivity of the network (density of links) is a global property (Giuliani, 2019). All these descriptors (and many others) are strictly intermingled across different organization layers. Thus, characteristic length inherits from the 'bottom' the information of the single node degree (higher degree nodes have an higher probability to enter into shortest paths). In turn, betweeness of a node (the number of shortest paths passing by a node, thus a microscopic feature of the network) inherits from the 'top' (mesoscopic level) the existence of clusters (modules) of nodes.

In this way, a node in between two different A, B clusters is traversed by all the shortest paths linking the A, B node pairs so scoring an high betweeness (Fig. 2).

Describing a system by network formalism implies a multi-level structural representation without the need of 'imposing' a particular bottom-up or top-down causative pattern.

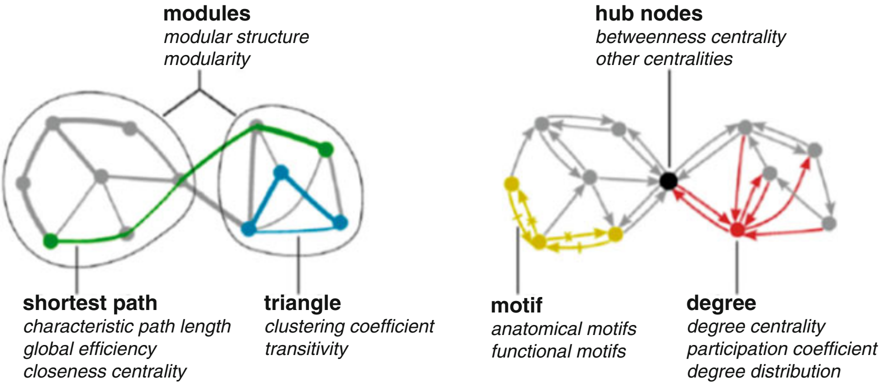


**Fig. 2** The figure reports schematically the most common graph invariants. Each index concentrates on a particular aspect of network wiring, shortest paths, participation coefficient and betweeness centrality are particularly important for describing fluxes across the network, clustering coefficient and modularity point to the existence of 'structural domain' within the network

## 4 Information Fluxes Within Networks

Proteins are the smallest objects that have all the features typical of complex systems, it is not without reason that the title of a seminal work on protein structure and dynamics (Frauenfelder & Wolynes, 1994) is '*Biomolecules: where the physics of complexity and simplicity meet'*.

Proteins 'sense' the environment, can acquire different stable state configurations, have an emergent behaviour not predictable from the accurate knowledge of their composition and perform complex 'actions' relevant for the system that host them. In addition the structural and compositional knowledge we have on protein molecules is order of magnitudes mere detailed and reliable than for any other complex system. This makes protein sciences a perfect playground for complexity studies.

Probably the most straightforward paradigm of information transfer through a network is the allosteric effect. Allostery is a neologism coming from Greek language, which has to do with the ability of proteins to transmit a signal from one site to another in response to environmental stimuli. The sensing (and consequent adaptation) of relevant information from the microenvironment is crucial for protein physiological role. This ability relates to the transmission of information across the protein molecule from a sensor (allosteric) site to the effector (binding) site (Di Paola & Giuliani, 2015). The protein molecule, hence, perceives ligand binding (or any other micro-environmental perturbation) at distance from the active site (where in turn the effective action takes place, e.g. where two small molecules are put together in order to catalyse their chemical reaction), and adapts its configuration accordingly. Thus, haemoglobin molecule senses at the allosteric site the partial

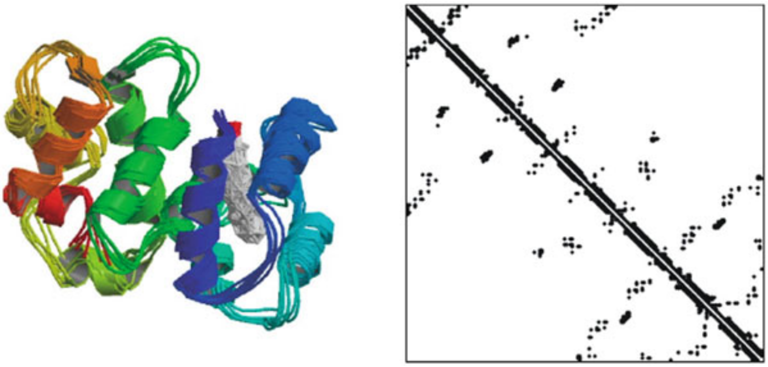


**Fig. 3** In this figure, the left panel the 3D structure of a small protein (recovering) follows the usual 'ribbon' style: the polypeptidic chain is represented in terms of contiguous segments of 'secondary structures' namely helices, random coils, and beta sheets. The right panel represents the same protein in terms of the adjacency matrix of the corresponding network (PCN = Protein Contact Network) whose nodes are the constituent amino-acids while the darkened pixels mark the relevant between amino-acid residues contacts (the unit values of Fig. 1)

pressure of oxygen (p[O2]): when p[O2] is high the affinity of haemoglobin for oxygen increases and the protein binds oxygen molecules at active site, on the contrary when p[O2] is low, affinity decreases and bound oxygen is released to cells. This process is crucial for life: in lungs there is a very high oxygen pressure and the haemoglobin present in red blood cells must catch oxygen molecules that in turn must be released in peripheral tissues (low p[O2]) so to make oxidative metabolism possible. How the protein molecule can discriminate such a relevant signal from the continuous perturbations of its structure coming from thermal noise and transmit the information at distance so to reach the active site.

To answer this question is useful to consider a protein molecule as a network (Fig. 3) having as nodes the aminoacid residues and as edges the intermolecular non-covalent bonds between residues generated by the 3D folding of the molecule. These networks are called Protein Contact Network (PCN) (Di Paola et al., 2013).

In Fig. 3 the aminoacid residues are ordered along the protein sequence from the left to the right in the X axis of the adjacency matrix and from the top to the bottom on Y. The 'trivial' contacts between aminoacids adjacent along the chain are not considered. This implies the scored contacts (links of the PCN) correspond to non-covalent intermolecular bonds putting different parts of the molecule into close contact (see Fig. 4, where a protein molecule is represented as a bracelet having aminoacid residues as pearls and PCN relevant contacts as red dashed lines).

In PCNs the shortest paths passing by the network edges mediate concerted motions and energy transmission upon stimulation of allosteric site (Di Paola & Giuliani, 2015). The topological metrics of shortest paths (minimum number of links separating two residues) is thus the actual metrics for signalling. Recently it was demonstrated (Poudel et al., 2020) that this purely topological metrics is

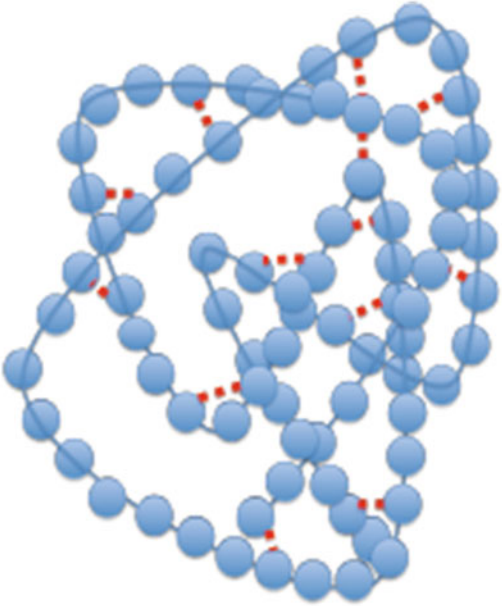


**Fig. 4** The blue line sequentially connecting the different aminoacid residues (pearls) corresponds to the covalent bond that generates the primary structure (sequence) of the macromolecule. In solution the protein molecule acquires its 'native' form by a folding process that generates its 3D structure responsible of its physiological role. The folding process puts aside residues otherwise distant along the sequence creating contacts (dashed red lines) among them. These contacts allow for a direct communication of the interacting aminoacid residues

coincident with the dynamical modes of protein molecule. This creates a spatio-temporal link of 'sustained modes' fulfilling the stable oscillation constraint we set for biodynamic interfaces. Thus we can say we are in presence of a 'fine tuned' grid deciding of the fate of external stimuli across the system.

The discrimination between relevant signals to be transmitted at distance without loss of information and non-informative perturbations to be dissipated without relevant changes in the 3D structure, relies upon two very important mesoscopic network descriptors: 'Guimera and Amaral' z and P indexes (Guimera & Amaral, 2005). The index z quantifies the number of contacts a given node (aminoacid residue in this case) has with other nodes of its own cluster (local contacts), while P scales with the number of edges linking the node to aminoacid residues pertaining to different clusters.

A perturbation affecting specifically an 'high P' node travels a long distance across the network passing by subsequent 'high P' nodes and arriving at destination supporting allosteric effects, on the contrary generic (noisy) thermal motion rapidly dissipates distributing across non-directional cycles thru intra-module motions.

High P nodes create a 'fast lane' for relevant information neatly separated by noise. This is exactly the role of biodynamic interfaces: some proteins, called multimeric, consist of distinct chains held together by intermolecular contacts. This is the case of haemoglobin made of four distinct polypetidic chains: the allosteric effect ends up into a different re-arrangement of the relative positions of the four chains that go back and forth between two different patterns (R and T for Relaxed and Tense) with high and low affinity for oxygen. The interface between these

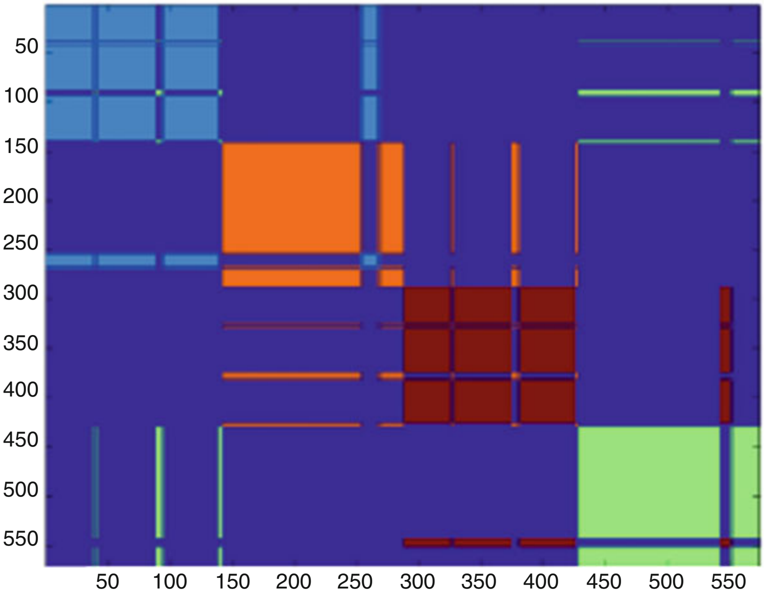


**Fig. 5** The figure reports the adjacency matrix of haemoglobin described by a colour code. The axes of the matrix reports the order of the residues along the chains (each chain corresponds to 150 residues), the dark blue corresponds to the lack of contacts, different colours correspond to the four chains

four chains is made of high P aminoacid residues that allow for the among chains concerted motions. Figure 5 gives a pictorial description of the situation.

From Fig. 5 it is evident the presence of 'displaced contacts' in the form of residues that, while pertaining to a given chain (module of the network) have the majority of their contacts with residues pertaining to different chains. These 'displaced contacts' are the long 'whiskers' contacting zones different from their own cluster (e.g. the pale blue line pertaining to the first chain (1–150)) that is in contact with the orange (second chain) module. These whiskers correspond to high P nodes that generate 'something in between' the interacting systems with a 'shared ontology' across the interacting systems (polypeptide chains).

Perturbations relevant for the allosteric effect (signals) enter the fast lane passing by high P residues and arriving at destination, on the contrary, not relevant (noisy) perturbations instead dissipate along futile within-module circuits. The presence of both fast (directional low loss) and slow (no-directional high loss) lanes of communication is shared by all natural networks (Kohestani et al., 2018) even if in protein molecules is much more evident than in other natural networks.

The discrimination between relevant and irrelevant stimuli is a form of 'meaning creation' by purely structural means that allow for a causative process embedded (and not imposed from the external) in the relational structure of the system at hand. This kind of causation makes obsolete the bottom-up top-down distinction and asks for a different explanation style in terms of ‘attractor-like’ dynamics spanning different layers of organization.

## References

Arora, M., Giuliani, A., & Curtin, P. (2020). Biodynamic interfaces are essential for human–environment interactions. *BioEssays, 42*(11), e2000017.

Demongeot, J., Aracena, J., Thuderoz, F., Baum, T. P., & Cohen, O. (2003). Genetic regulation networks: Circuits, regulons and attractors. *Comptes Rendus Biologies, 326*(2), 171–188.

Di Paola, L., De Ruvo, M., Paci, P., Santoni, D., & Giuliani, A. (2013). Protein contact networks: An emerging paradigm in chemistry. *Chemical Reviews, 113*(3), 1598–1613.

Di Paola, L., & Giuliani, A. (2015). Protein contact network topology: A natural language for allostery. *Current Opinion in Structural Biology, 31*, 43–48.

Fankhauser, G. (1945). Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. *Journal of Experimental Zoology, 100*(3), 445–455.

Frauenfelder, H., & Wolynes, P. G. (1994). Biomolecules: Where the physics of complexity and simplicity meet. *Physics Today, 47*(2), 58–64.

Giuliani, A. (2019). The actual status of quantitative approaches in biology: Problems and perspectives. *Rend Mat Appl, 40*(7), 165–176.

Guimera, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature, 433*(7028), 895–900.

Kohestani, H., Totonkuban, M., Di Paola, L., Todde, V., & Giuliani, A. (2018). The basic principles of topology-dynamics relations in networks: An empirical approach. *Physica A: Statistical Mechanics and its Applications, 508*, 584–594.

Laughlin, R. B., Pines, D., Schmalian, J., Stojković, B. P., & Wolynes, P. (2000). The middle way. *Proceedings of the National Academy of Sciences, 97*(1), 32–37.

Mendonça, V., Madeira, C., Dias, M., Vermandele, F., Archambault, P., Dissanayake, A., & Vinagre, C. (2018). What’s in a tide pool? Just as much food web network complexity as in large open ecosystems. *PLoS One, 13*(7), e0200066.

Mickulecki, D. (2001). Network thermodynamics and complexity: A transition to relational systems theory. *Computers & Chemistry, 25*, 369–391.

Nicosia, V., De Domenico, M., & Latora, V. (2014). Characteristic exponents of complex networks. *Europhysics Letters, 106*(5), 58005.

Pascual, M., & Levin, S. A. (1999). From individuals to population densities: Searching for the intermediate scale of nontrivial determinism. *Ecology, 80*(7), 2225–2236.

Poudel, H., Reid, K. M., Yamato, T., & Leitner, D. M. (2020). Energy transfer across nonpolar and polar contacts in proteins: Role of contact fluctuations. *The Journal of Physical Chemistry B, 124*(44), 9852–9861.

Sporns, O. (2018). Graph theory methods: Applications in brain networks. *Dialogues in Clinical Neuroscience, 20*(2), 111.

Weaver, W. (1948). Science and complexity. *American Scientist, 36*, 536.

# From Predictability to the Theories of Change

**Ignazio Licata**

**Abstract** The study of processes and change in systems is a requirement for Theoretical Physics after the development of complexity and emergence theories. This matter is far beyond the ideal models—centered on predictability—that Mathematical Physics usually deals with. The strongly interdisciplinary and systemic issue of a Theories of Change implies a careful reconsidering of the essential features of the relationships between the observer/model-builder and the system under consideration. We delineate here such relationship as the meta-theoretical step wherein it is possible to give the first and partial collocation to the wide class of not ideal models, and to evaluate the effective forecasting possibilities of Big Data.

**Keywords** Complexity · Emergence · Ideal and not-ideal models · Predictability · Change · Big Data

## 1 Introduction: The Climbing of Mount Epomeo

Mount Epomeo (789 m, as Wikipedia says) is the highest mount of the isle of Ischia. The legend tells it is one of the four entrances to Agharta. As for me, it surely was the occasion for an extraordinary encounter. It was a very hot end of May in 1991, Giuseppe Arcidiacono—the great mathematician and cosmologist of Projective Relativity I first met through an intense epistolary exchange and, later, during my military service in L'Aquila—invited me to a Conference organized by University of Perugia and the Istituto Filosofico of Naples. I had the chance to discuss with

---

I. Licata (✉)
ISEM, Institute for Scientific Methodology, Palermo, Italy

School of Advanced International Studies on Theoretical and Non Linear Methodologies of Physics, Bari, Italy

International Institute for Applicable Mathematics & Information Sciences (IIAMIS), B.M. Birla Science Centre, Hyderabad, India
e-mail: ignazio.licata3@gmail.com

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_6

many interesting people there, such as J. Barbour and L. Kostro. Anyway, getting acquainted with Eliano Pessa has modified and broadened enormously my Physics and has endowed me with a vision for Systemics. I had recently published some papers (I find them quite naive now) about lattice space-time on Plank scale. The lattice worked as a constraint for a set of oscillators which produced the observable particles . Just to give an idea to the contemporary reader, it was an idea similar to what 't Hooft has been recently developing, far more elegantly than me, indeed (Licata, 1991, 2003; 't Hooft, 2016). What he first said was: "I read your work. That's one of the best thing I have seen this year". We had an intense conversation and we keep on talking in the afternoon, in a quite unusual way for me. Eliano was an expert alpinist who had already climbed the most challenging mountains on the Planet, such as the Nanga Parabat, and yet he couldn't resist to the temptation of ascending, as short as it may be. So, he proposed me to climb Monte Epomeo, while climbing each of us told about his life. Reaching to the top was not an event worthy to be recorded in the Alpinism annals, but it was the celebration of our friendship birth as well as an enduring collaboration made of meetings and periods at high epistolary intensity. Eliano Pessa had his Ph.D. in Physics with Bruno Touschek, had collaborated with the mathematician Bruno Rizzi for a long time, had found a new interest in Cosmology with Giuseppe Arcidiacono, and, thanks to a formidable series of works on neural nets, was going to became the first theoretical physicist Full Professor in Psychology at The University of Pavia.

These few notes clearly give the idea of a vision far beyond the slightly narrow horizons of a traditional physicist, usually connected to the rhetoric of infinitely small or big.

In that period, Eliano was working on some problems related to the synergetic approach (Haken, 2012). The Haken Theory, on the strength of the work on collective behaviors in laser and on phase transitions, suggested that in many different contexts—Biology, Sociology. Etc.—it was possible to find situations where few parameters "took command" of all the other variables into play, which thing provided a more elegant description and, if not (asymptotic) predictability, a kind of understanding of the "possible destinies" of the system. It has to be said that the Haken conjecture gets a strong and immediate meaning: if there weren't exist the variables that "mediate" between the components or the agents (microscopic level) and the possible global outcomes (macroscopic), it would be nearly impossible distinguish between information and entropy because the non-cooperative aspects of the system would prevail. Actually, there wouldn't be any authentic emergence. That is just a well-founded conjecture. The real problem lies in questioning if it is possible to classify systems depending on how they produce information. According to Systemics general terms, there exists the Theorem of R. Shaw (1981), one of the members of the legendary chaos collective, which establishes a relationship between the phase space of a system—whose coordinates are the variables fixing its behavior—and some features of the equations describing them (non-linearity degree, constraints and so on). Accordingly, we can distinguish between: (a) Information-conserving Systems where the principle of energy conservation is valid; (b) Information-amplifying polynomial Systems, where information increases along time until a certain state of equilibrium. All the self-organizing models belong to this category, such as the dissipative structure by I. Prigogine. It has long be debated whether such models—and catastrophes, their geometrical counterpart—could be the universal key of morphogenesis, but some fundamental theorems, among which those by R. Landauer and R. Fox, and then the N. Koppell and D. Ruelle, have shown that there are very precise limits to complexity for the configurations that these systems can reach. Finally: (c) Information-exponential amplifying Systems, which chaotic systems belong to. The Theorem of Shaw is a very cornerstone because it introduces a way to look at systems no more centered on non-linear complications, but as classes of complexity of systems. It is patent that the "interesting" systems can be found in a narrow range between the moderately ordered of systems (b) and the savage proceeding towards chaos of systems (c). That poses extremely fascinating questions about the relationships—for example—between Physics, Biology and Cognitive Sciences (Licata, 2018; Pessa, 2008; Vitiello, 2001; Freeman & Vitiello, 2008), and tips the epistemic scales from the model as a mirror of an "objective" world to the specific conditions in which it is actually possible to apply it to something we individuate as a system. The awareness of a conditional and never banal correspondence between models and systems is the basis of the recent intense interdisciplinary crossing—models can "migrate"!—, but it is also a Copernican Revolution in the way we look at the constructive activity of science.

## 2 The Scientific Explanation and the Causality-Determinism-Predictability Triad

Predictability is still considered as a crucial ingredient in scientific explanations. Likely, this close association depends on the historical fortunes of Determinism, a peculiar trait for a lot of physical theories from Classical Mechanics to Relativity up to Quantum Mechanics, anyway the predictability into play changed its features and range at each step. As it is known, a physical system is described by deterministic laws when, given the dynamic laws and the initial conditions, the "mathematical crank" (differential equations) univocally fixes a state of the system in a future or past instant. The successes of Rational Mechanics confirms the philosophical triumph of the Laplace's demon (Licata, 2015) and for a long time the huge complications linked—for example—to the three-body general problem were considered as computational problems or just a matter of inexact data. It will be a 1887 work by Poincaré to clarify that the problem actually pertains to a new typology, so prefiguring the modern theory of dynamical systems based on chaos and non-linearity. In this case, the sensible depending on the initial conditions limits very quickly the predictability in a range strictly connected to the system's non-linearity (Lyapunov time). In addition, it can be demonstrated that a non-linear system is an information amplifier—it can, for example, "inglobate" a fluctuation—

thus the connection between determinism and predictability completely falls down (Cencini et al., 2009; Licata, 2008a). The qualitative theory of dynamical systems has developed since the 30s in order to solve these situations by proposing a concept of predictability no more based on the detail of the specific trajectories, but on the global behavior of the system, i.e. stability and asymptotic states (Kutznetsov, 2004). We find this situation also in General Relativity, the difference—and some subtler problems more, see the specialized bibliography—lies in the fact that there is no background, but the space-time itself is generated by the initial data and the evolutionary equations (Barrow, 1982).

The case of Quantum Mechanics is a completely different one. Actually, if the evolutionary dynamics U (for example, the Schrödinger's equation) are strictly deterministic, any predictability on a single event (R, reduction of the state vector) is impossible for principle reasons linked to the very nature of quantum systems. The whole debate on foundations can be seen as the different attempts to find a synthesis between U and R, or, at least, the "ontological" elimination of one of the two sides.

Despite its "erosion" within the hard fabric of theoretical physics, predictability seems to maintain a special position. Actually, it is considered that determinism distills the scientific features of causality, a long debated notion in philosophy. Clearly such idea does not stand to a careful examination. The possibility to connect two events in a temporal framework (for example in the light cone structure in relativity) does not guarantee by itself a causal relation (Mumford & Anjum, 2014; Illari & Russo, 2014). Moreover, the local features of Quantum Mechanics would make the question much more problematic (Näger, 2016; Ringbauer et al., 2016; Pegg, 2006; Popescu & Rohrlich, 1998).

Thus, it is more natural to connect the notion of "cause" with the global structure of a scientific explanation, considered as the configuration of theories and models providing a picture of how things work! Predictability is rather a feature of some classes of models. We can easily realize they are different things by a simple example. Just consider the classical double slit experiment for electrons by C. J. Davisson and L.H. Germer in 1927, which R. Feynman correctly considered as the archetypal QM experiment (Feynman, 1985). Let's imagine we could collect all the data of the electron impacts on the screen and process them statistically. In the end, we can get a very good probabilistic evaluation of the areas where we can find an electron, no more, no less than we use wave-function. Anyway, such prediction does not explain the phenomenon. In order to have an exhaustive picture of the situation we have to turn to Schrödinger equation; moreover, our statistics cannot be exported to other phenomena, the Schrödinger equation, instead, is an explanation connecting scales, objects and dynamics for non-relativistic quantum systems.

The separation between explanation and predictability becomes far more evident in the study of complex and emergent systems. Without going into technical details, we know that – in an emergent process – redefining the system's internal structure and modifying the relationship with the environment leads to the emerging of new properties which, generally, cannot be ascribable to the level of its constituents, in a way that remind us the universal aspects of phase transitions (Licata, 2008b, c). It can imply different levels of description, and the manifestation of some forms of coherence on some levels and not in other ones. In general, intrinsic emergence is connected to the appearing of properties or phenomena compatible with the models which describe those phenomena, but—in principle—they are unpredictable considering that the models, being conditions the same, admit more different behaviors and the new properties or phenomena can modify irreversibly the nature itself of the system (Licata 2010; Ryan 2007; Anderson & Stein, 1985; Ronald et al., 1999; Goldstein, 1999; Pessa, 2006, 2008; Bedau & Humphreys, 2008). So, the problem is to understand the physical and mathematical reasons why some classes of models admit a strong predictive apparatus as a consequence of their descriptions and some other not, and—in this case—what evaluations and interventions are possible.

## 3 Ideal Models, Phase Space and Boundary Conditions

Predictability is a mathematical feature of a model and corresponds, all in all, to a series of physical conditions satisfied by the system under consideration. We will focus here on these ones by taking into consideration some essential points that constraint the building of the so-called ideal models, susceptible either of closed analytical solutions or, anyway, of approximations that assure a broad predictability, if not a local one, at least on its asymptotic states. Such wide class of models cover a quite big theoretical range encompassing from Newton Mechanics to Quantum Field Theory (QFT).

Differently from what is often stated, reductionism—i.e. the exclusive attention for constituents and explicative downward arrows—is not an essential prerequisite to have predictability. This idea derives from classical trajectories, but it is easy to realize that it is not so just by thinking to a perfect gas in a jar. If we open the jar, the gas will diffuse in the environment; this will fixes a time arrow based on entropy. In spite of it, the level of gas particles is always ruled by a reversible dynamics unchanging if we “rewind the film”: in fact, it is ruled by the Boltzmann Stosszahlansatz. Anyway, the evolution of the system is possible by using a diffusion equation applied to the whole statistic ensemble. This exemplary case demonstrates that it is not so much the request of reductionism the key for predictability as the particular connection interplay between different micro, meso and macro levels. In this case, the physics and mathematics of the phenomenon is fully fixed by the mirable bridging by statistical physics between Newton Mechanics and Thermodynamics (Cercignani, 2006).

The case of molecular chaos is interesting because we can introduce some important elements more. The “identity” of particles and forces into play is always well defined and fixed, moreover, temporal reversibility at microscopic level is conserved, and the environment is just a very elementary and “passive” scenario. “The first” and “the after” of the phenomenon takes place by modifying the boundary conditions of the gas when the jar is opened. One of the reasons which contributed in creating a reductionist-deterministic mythology has been the fact of neglecting the decisive importance of the boundary conditions. If we forget them—and the environment influence as well—, physical laws turn into good for all season mere algorithms and, above all, indifferent to the level under consideration. That is not so! Our example reminds us that a level of analysis does not guarantee its portability to other levels. For molecular chaos, we can, at least, connect macro and micro thanks to the statistical interpretation of thermodynamic quantities, it is impossible for most of the real interesting complex systems. On the contrary, emergence processes suggest that the creation of new models and the "convergence" of different approaches into a "super-model" is not for sure. The complementary between laws and conditions is well described by D. Bohm (1957) when he states that a physical law is a frame of possibilities, it will be the chance (in etymological sense) to decide which phenomenon will take place. It could sound obvious for quantum systems and inappropriate for an aseptic world of Newtonian balls, but it is not so for systems where noise, fluctuations, non-linearity and the strong variations ruling the system-environment relations make that frame greatly sensible to chance with modifications of boundary conditions. Complex systems can be defined as systems where the variation of boundary conditions is more important than laws, because they modify the role of the laws involved—by a drastic complexifying of the phase space—and the nature itself of the "objects" into play (Licata, 2012, 2015; Licata & Minati, 2016).

Let us now consider with a little more attention the powerful theorem on the non-linear systems' behavior from the informational viewpoint by Shaw (1981). Suffice it to remind that the system's volume in the phase space (i.e. the space of the behaviors in relation to a set of observables) modify its information content at a rate connected to the peculiar type of non-linearity of the system. Such theorem can be thus considered as a non-linear filiation of the Liouville Theorem. A critical notion to focus our attention on is the system's volume. Even without any mathematical lingo, it easy to see that it is closely connected to boundary conditions. In other words, it is supposed that the system/environment relationships can be clearly described by a handful of fixed parameters, it makes possible monitoring the information exchange. In spite of its being a limiting condition (just think to Maturana and Varela's system/environment coupling for living systems, Maturana & Varela, 1980), many interesting systems are included in this classification. Information amplifying systems are particularly interesting for the topic of predictability. A first category are the polynomial amplifiers, which include the well-known dissipative structures (Nicolis & Prigogine, 1989); these systems evolve towards self-organization states by means of a balancing feedback between outgoing entropy and ingoing energy. Such approaches has been welcomed with great enthusiasm, at first they were considered able to provide a general framework for the order out of chaos program, soon, many limiting theorems (Kopell & Ruelle, 1986) have showed that the complexity of these structures is rather poor. The expectations for H. Haken Synergetics (Haken, 2012) met the same destiny. Synergetics treats the emergence of some order parameters on mesoscopic regime that lead the system toward more organized states, by taking control of the microscopic variables. In spite of the interesting analogies that such scenarios have stimulated for systems very far from physics or chemistry, their importance in biology is limited and purely metaphoric in socioeconomic fields.

What is worthy noticing is that in all the above-mentioned cases, the hypotheses on the system's openness are such that there can be applied some mathematical techniques corresponding to a careful mapping of information in the phase space. With dissipative and synergetic systems, non-linearity allows a simplified situation when passing from the microscopic to the mesoscopic, by merging many down level disordered dynamics into an ordered scenario. All that guarantees high predictability. Also in the case of exponential amplifier systems—the chaotic system in strict sense—, there are many different techniques to monitor the transition to chaos and the asymptotic states (at least for low dimensional systems). We are there at the extreme borders of predictability, whose essential condition seems to be the possibility to get models able to follow the system's phase space. Such condition is equivalent to apply a series of hypotheses related to symmetries, balancing laws and constraints about the system-environment relationship to the system; these hypotheses turn into the possibility to develop an analytical treatment with strong predictive features (Vakulenko, 2013).

It has also been suggested to use the Quantum Field Theory (QFT) formalism as a general framework to study formally the dynamics of phase transition with intrinsic emergence. In these systems, emergence is introduced as symmetry breaking and there new types of long-range correlations are created. Despite some interest-worthy cases, such as the Quantum Brain by G. Vitiello, everyone agrees that these approaches are quite far from the actual complexity of a biological system. (Pessa, 2002; Vitiello, 2002).

After having thoroughly explored the possible links between QFT as a framework of complexity, in recent years Eliano Pessa strongly expressed the idea that QFT was too "narrow" for the study of complex systems:

> *Unfortunately, the intense research activity on emergence has not yet produced a universally agreed definition of what is meant by 'emergence'. There is however a widespread consensus regarding the general aspects that should characterize any form of emergence (the Goldstein's work, 1999 is very useful in this regard). A synthetic list of these aspects includes:*
>
> *a) dependence on an observer; this means that emergence is not an objective property or event in Nature, but it is linked to the surprise an observer (equipped with tools concepts, theories, beliefs, intentions, purposes and so on) feels when - placed in front of a system - there sudden appear properties not attributable to any observable efficient cause in the system itself;*
>
> *b) the existence of different levels of description; this is equivalent to say that emergence can reveal itself only by adopting an appropriate level of description.*
>
> *c) the appearance (or disappearance) of some form of coherence; without going now into the very difficult question of what exactly is meant by 'coherence', here we limit ourselves to point out that in all emergence processes so far considered as such (not only in the physical sphere, but also in biological and social one) we always managed to identify some kind of consistency that was acquired (or lost).*
>
> *(...)*
>
> *a grounded- both philosophically and scientifically - theory of emergence processes - is yet to come. The TTPT(Traditional Theory of Phase Transitions) certainly represented a starting point to build such a theory. However, despite the mathematical sophistication of*

> *the tools employed, the TTTF remains too low a theory, inadequate not only to account for the processes of emergence in the biological, psychological, social and economic fields, but also to explain many processes in Physics. Among the possible remedies for this inadequacy, we have tried to emphasize the introduction of adequate theories of system-environment interactions.*
>
> *These arguments have shown us how this path, still in the initial stages, is fraught with both technical and conceptual difficulties. Moreover, it does not sound strange: for centuries science and philosophy have just taken into consideration very simple systems where it was not necessary, by hypothesis, to worry much about the environment they were immersed in. Therefore, they never bothered to define what environment was, what it was like and how to distinguish it from the systems immersed in it and how to model it* (Pessa, 2012).

## 4 Recalcitrant Systems and Configurational Variables

There is thus a problem in fixing some common features to recalcitrant systems. These ones resist to mathematical formalization and, anyway, do not offer a firm anchorage for the traditional notion of predictability.

The concept itself of equilibrium state does not make sense for these phenomena. Such systems are continuously evolving, often very fast, and cannot be "zipped" within a model because their nature is essentially that of a process. In these cases, neither the system's nature nor the environment can be regarded as fixed and characterized by few parameters; the attention, instead, has to be focused to the kind of coupling in order to put into evidence what aspects can influence the evolutionary characteristics which continuously undergo a multi-level plurality of meta-stable situations of adaptation. Strictly speaking, we should not to speak about a system-environment distinction, but about classes of events. In particular, lately, there has been an intense activity on networks—a strategic architecture in nature and artificial system too—that clearly shows another important difference: the one between dynamics and history. In networks, the dynamic behaviors can lead to the disappearing/creating or reinforcing of some hubs so modifying completely how the system manages information (Barabasi, 2018; Boccaletti et al., 2006; Costa et al., 2007). In this situation, some frozen components can come out; they mark the system's history in the form of stratified constraints which act on temporal scales much longer than the ones of other nodes. In a much more radical way than the case of the gas in the jar, the system's global history develops on a different level from the individual dynamics. Moreover, such transitions do not take place near a critical point, as it happens in the traditional phase transition theory, and this makes investigating these situations even farther from the ideal models. The changes linked to complex structures of internal constraints tend to make the system autonomous and its relation with environment is highly selective, so introducing a semantic dimension in the informational flux; this is, maybe, the current strongest limit for the mathematical modeling, which allow some sort of predictability. It is clear that in these situations, such as the conservation of energy (as well as the microscopic syntactic information) is naturally compatible, but it does not say anything about the global semantic choices. Actually, it is merely true also in "simple" systems:

a physical description of the energy dissipated by a computer hardware does not indicate us anything about the characteristics and the complexity of its software. A strong evidence of the global importance of semantics has came from the study of biological neurons, it has been found out that the electrical activity of the same neuron crucially depends on the global functions performed by the neuron network it is immersed in (Sahasranamam et al., 2016).

We can realize, even from these few lines, that the majority of interesting complex systems—for example, the socio-economic ones—are recalcitrant to ideal modeling. This is the big challenge of the theories of change, which is not referred to a formal structure such as the T theory and M model—axiomatic or semi-axiomatic—of theoretical physics, but to a set of multi-model strategies focused on the observative and computational inquiring of the events under consideration. It is not as much predictability into play, but rather the understanding of the change factors and—above all—the intervention on an actual process.

The good news is that the extremely recalcitrant systems—i.e. living within pure disorder and perfect casualty—, are very rare and, anyway, are not so interesting to study. An authentic casual sequence, according to Kolmogorov, is an infinite sequence that cannot be zipped in a string shorter than the sequence itself, we can only observe its evolution step by step. Actually, stating that for any finite sequence it is always possible to find a rule or interpolation connecting them does not solve any empiric problem. What is really interesting in a network of event is to individuate information, at least, as meaningful configurations for an observer. The 300-years-old implicit assumption of theoretical physics is that our descriptions correspond to something in the World out there, it is surely true, but we have not to forget we are the ones who build such descriptions. We can do it because the systems ruled by ideal systems are simple. The challenge comes out for the systems suspended between order and disorder, where "at the edge of chaos, the boundaries of change fluctuate endlessly between a sluggish status quo and the anarchy of perpetual destruction" (Cohen, 1997). In these systems, to bridge the microscopic and the macroscopic levels in a simple way is almost always impossible, but there emerge configurations always remodeling the system and make possible to identify some mesoscopic variables which mediate between the two levels and characterize the metastable state of the change. These are not "observables" in a traditional sense, but a choice between the patterns an observer individuates as the expressions of the change. Here, we will focus neither on the available analytical tools nor on the formidable problems of quasi-ergodicity (Bertuglia & Vaio, 2005; Moore, 2015), in the end, we will try instead to fix some conceptual aspects the study of these systems offers.

## 5 Metastructures and Big Data Forecasting

Metastructures are a general framework for studying change. The question they come from equally belongs to cognitive sciences (observer) and Physics (observed):

how can a change in a process be detected? Since the very beginning an observer is called to choice what to observe and how to do it, by trying to spot the interesting characteristics which occur quite regularly. We are speaking about the intuitive emergence according to the famous classification by Crutchfield (1994), anyway, considering we are dealing with process that are quite far from ideal models, it is the only emergence we can detect, because there are not any available mathematical structures to study the dissipative forms or intrinsic emergence. So, investigating meta-structures is a research not only focused on the emergence features and the change in a system in "objective" sense, but also on the dynamic relationship between an observer and an observed system. For the technical aspects, see references (Licata & Minati, 2013; Minati & Licata, 2013).

The procedure is based on building one or more mesoscopic vectors whose values are chosen by an observer. In the "pedagogic" flock example the typical values can be: velocity, directions, habits, distance—maximum and minimum—between two "constituents", the instantaneous value in a time interval of the flock surface of volume, and so on. In this way, it is possible to get information from the system to individuate and detect change, its phases, sequences and modalities; for example, what elements play the same role at different instants or different roles at the same instant. In other words, the mesoscopic vector is an active "grid" that probes the system by means of computational procedures in search of meaningful signals for the observer/model builder. Such scenario offers some reflections about Big Data.

It is true that data did not become big overnight, in Jeff Jonas words (Jonas, 2012), and we can add that science has always strongly needed data not only to build theories, but to correct them, delimit them and, finally, to make them operative. The Navier Stokes equations would just be a conceptual scenario for meteorologists, if it weren't for a thick network of sensors, and it surely does not depend on their mathematical status of Millennium Clay Problem! Anyway, the debate about Big Data has become hot about epistemological and ethical problems just lately (Cardon, 2015; Mayer-Schönberger & Cukier, 2013). The Chris Anderson provocative paper for Wired (Anderson, 2008) has triggered many relevant answers among which we cannot avoid reminding the Calude and Longo one (Calude & Longo, 2017): in general, and for very good mathematical reasons, data don't speak for themselves, most of the correlation are spurious. It happens only if we look at correlations as "objective", and—above all—if we limit our research to correlations! If we accept that this is new brand field, and that data have to be investigated to speak then we will be in a world of possibilities similar to what Ermanno Bencivenga described: So there you have it. Big Data enthusiasts are (unwittingly) advocating a new definition of what it is to know. Their agenda is (unwittingly) semantical. Except that it is not worked out, and any attempt at developing it in the semantical terms that have been current (and antagonistic) for the past two millennia is hopeless (Bencivenga, 2017).

If we try to look beyond the simple correlations and consider Big Data as something more than a passive object of multivariate statistics, we can see an extraordinary number of patterns with formidable theoretical implications, such as the power law, non-Gaussian behaviors, scenery of criticality, fat fractals, and so on (for a general textbook on these techniques: Sethna, 2006; for a class of exemplary problems: Buldyrev et al., 2010). It does not mean that the processes under consideration cannot be included within the old theoretical boundaries; we have, instead, to do theory in progress, to use the historical baggage within a new style of work consisting in testing heuristic hypotheses following the process, according to an abductive line.

The current classification of Big Data in the so-called 6Vs model (volume, velocity, variety, veridicity, variability and value) is still too tied to the quantitative data flux and an objectivist conception. It's advisable, and greatly plausible, that the structures analytics (descriptive/predicative) more and more point towards the search of the secret life of Big Data, in order to do it we have to realize that analytics directly express what meaning the process has for the agent observing it. It is not a matter of building up a robot-scientist, dear to the old AI, but to get open to the idea that in the same way as high speed and extremely small have changed theoretical physics, the extreme complexity will modify our concept of Physics in an even more radical way.

## 6 Conclusions

In the history of theoretical physics, predictability has established itself within deterministic explanations, then it has extended, in its different forms, to the class of ideal model, which allow an univocal description of a system. In complex systems at high emergence, univocity breaks down and the descriptions become plural, that is an extreme limit for predictability in strict sense. Many strategies and tools for forecasting a process have been grown, they are connected to the meaning that a class of events have for the observer or agent. In many complex systems it is not just an abstract forecasting to solve the problem, but the actual awareness about the possibilities to guess on. All that strengthens the Bruno de Finetti (1906–1985) intuitions: *Knowing how things will go, as if they were occurring on their own (...)It is a problem of decision, not prevision ( . . . )...- Shouldn't the exceptional dimension of our empiric world lead to a completely new awareness? How can we expect that logic chains end with certainties, just like a good ole syllogism?* (de Finetti, 1968, 1972).

## References

Anderson, C. (2008, June), The end of theory: The data deluge makes the scientific method obsolete. *Wired*.

Anderson, P. W., & Stein, D. L. (1985). Broken symmetry, emergent properties, dissipative structures, life. Are they related? In F. E. Yates (Ed.), *Self organizing systems: The emergence of order* (pp. 445–457). Plenum Press.

Barabasi, A. L. (2018). *Network science*. Cambridge University Press.

Barrow, J. (1982). Chaotic behaviour in general relativity. *Physics Reports, 85*(1), 1–49.

Bedau, M. A., & Humphreys, P. (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science*. MIT Press.

Bencivenga, E. (2017). Big data and transcendental philosophy. *The Philosophical Forum, 48*(2), 135–142.

Bertuglia, C. S., & Vaio, F. (2005). *Nonlinearity, chaos, and complexity: The dynamics of natural and social systems*. Oxford University Press.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*, 175–308.

Bohm, D. (1957). *Causality and chance in modern physics*. University of Pennsylvania Press.

Buldyrev, S., Parshani, R., Paul, G., Stanley, E. H., & Havlin, S. V. (2010). Catastrophic cascade of failures in interdependent networks. *Nature, 464*, 1025–1028.

Calude, C. S., Longo, G. (2017). The Deluge of Spurious Correlations in Big Data, Found. *Science, 22*(3), 595–612

Cardon, D. (2015). *A quoi rêvent les algorithmes*. Nos vies à l'heure.

Cencini, M., Cecconi, F., & Vulpiani, A. (2009). *Chaos. From simple models to complex systems*. World Scientific.

Cercignani, C. (2006). *Ludwig Boltzmann: The man who trusted atoms*. Cambridge University Press.

Cohen, B. (1997). *The edge of chaos. Financial booms, bubbles, crashes and chaos*. Wiley.

Costa, L. F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics, 56*, 167–242.

Crutchfield, J. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D, 75*(1), 11–54.

de Finetti, B. (1968). Riflessioni sul futuro. *Civiltà delle Macchine, 16*(3), 82.

de Finetti, B. (1972). Atti del convegno "Tecnologia e problema ecologico" Roma 1972. *Civiltà delle Macchine, 20*, 3–4.

Feynman, R. P. (1985). *QED: The strange theory of light and matter*. Princeton University Press.

Freeman, W., & Vitiello, G. (2008). Dissipation, spontaneous breakdown of symmetry and brain dynamics. *Journal of Physics A: Mathematical and Theoretical, 41*, 304042.

Goldstein, J. (1999). Emergence as a construct: History and issues. *Emergence, 1*, 49–72.

Haken, H. (2012). *Synergetics: Introduction and advanced topics*. Springer.

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. Oxford University Press.

Jonas, J. (2012). Interview on data protection & law policy newsletter. https://jeffjonas.typepad.com/jeff_jonas/2012/04/big-data-qa-for-thedata-protection-law-and-policy-newsletter/comments/.

Kopell, N., & Ruelle, D. (1986). Bounds on complexity in reaction-diffusion systems. *SIAM Journal on Applied Mathematics, 46*(1), 68–80.

Kutznetsov, Y. A. (2004). *Elements of applied bifurcation theory*. Springer.

Licata, I. (1991). Minkowski space-time and dirac vacuum as ultrareferential fundamental frame. *Hadronic Journal, 14*, 3.

Licata, I. (2003). The big computer. Complexity and computability in physical universe. In V. Benci et al. (Eds.), *Determinism, holism and complexity*. Kluwer.

Licata, I. (2008a). Emergence and computation at the edge of classical and quantum systems. In I. Licata & A. Sakaji (Eds.), *Physics of emergence and organization* (pp. 1–25). World Scientific.

Licata, I. (2008b). Physics and logical openness in cognitive models. *Epistemologia, 2*, 177–191.

Licata, I. (2008c). *La Logica Aperta della Mente*. Codice Edizioni.

Licata, I. (2010). Almost-anywhere theories: Reductionism and universality of emergence. *Complexity, 16*(5), 11–19.

Licata, I. (2012). Seeing by models: Vision as adaptive epistemology. In G. Minati, M. Abram, & E. Pessa (Eds.), *Methods, models, simulations and approaches towards a general theory of change* (pp. 385–400). World Scientific.

Licata, I. (2015). *I gatti di Wiener. Riflessioni sistemiche sulla complessità*. Bonanno Editore.

Licata, I. (2018). *Complessità. Un'Introduzione semplice*. Di Renzo.

Licata, I., & Minati, G. (2013). Emergence as mesoscopic coherence. *Systems, 1*(4), 50–65.

Licata, I., & Minati, G. (2016). Emergence, computation and the freedom degree loss information principle in complex systems. *Foundation of Science, 22*(4), 863–881.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition*. Reidel.

Mayer-Schönberger, V., & Cukier, K. N. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Minati, G., Licata, I., & Pessa, E. (2013). Meta-structures: The search of coherence in collective behaviours (without physics). *Electronic Proceedings in Theoretical Computer Science, 130*, 35–42.

Moore, C. C. (2015). Ergodic theorem, ergodic theory, and statistical mechanics. *Proceedings of the National Academy of Sciences of the United States of America, 112*(7), 1907–1911.

Mumford, S., & Anjum, R. L. (2014). *Causation: A very short introduction*. Oxford University Press.

Näger, P. M. (2016). The causal problem of entanglement. *Synthese, 193*(4), 1127–1155.

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity: An introduction*. Freeman.

Pegg, D. T. (2006). Causality in quantum mechanics. *Physics Letters A, 349*(6), 411–414.

Pessa, E. (2002). What is emergence. In G. Minati & E. Pessa (Eds.), *Emergence in complex cognitive, social and biological systems*. Kluwer.

Pessa, E. (2006). Physical and biological emergence: Are they different? In G. Minati, E. Pessa, & M. Abram (Eds.), *Systemics of emergence. Research and development* (pp. 355–374). Springer.

Pessa, E. (2008). Phase transitions in biological matter. In I. Licata & A. Sakaji (Eds.), *Physics of emergence and organization* (pp. 165–228). World Scientific.

Pessa, E. (2012). Processi di auto-organizzazione e interazione sistemi-ambiente. *Rivista di Filosofia Neo-Scolastica, 4*, 639–659.

Popescu, S., & Rohrlich, D. (1998). Causality and nonlocality as axioms for quantum mechanics. In G. Hunter, S. Jeffers, & J. P. Vigier (Eds.), *Causality and locality in modern physics. Proceedings of a Symposium in Honour of Jean-Pierre Vigier* (pp. 383–389). Kluwer.

Ringbauer, M., Giarmatzi, C., Chaves, R., Costa, F., White, A. G., & Fedrizzi, A. (2016). Experimental test of nonlocal causality. *Science Advances, 2*(8), e1600162.

Ronald, E. M. A., Sipper, M., & Capcarrère, M. S. (1999). Design, observation, surprise! A test of emergence. *Artificial Life, 5*, 225–239.

Ryan, A. J. (2007). Emergence is coupled to scope, not level. *Complexity, 13*(1), 67–77.

Sahasranamam, A., Vlachos, I., Aertsen, A., & Kumar, A. (2016). Dynamical state of the network determines the efficacy of single neuron properties in shaping the network activity. *Nature, Scientific Reports, 6*, 26029.

Sethna, J. P. (2006). *Statistical mechanics: Entropy, order parameters and complexity*. Oxford University Press.

Shaw, S. R. (1981). Strange attractors, chaotic behavior, and information flow. *Zeitschrift für Naturforschung A, 36*(1), 80–112.

't Hooft, G. (2016). *The cellular automaton interpretation of quantum mechanics*. Springer.

Vakulenko, S. (2013). *Complexity and evolution of dissipative systems: An analytical approach*. Walter de Gruyter.

Vitiello, G. (2001). *My double unveiled*. John Benjamins.

Vitiello, G. (2002). Quantum field theory and systems theory. In G. Minati & E. Pessa (Eds.), *Emergence in complex cognitive, social and biological systems*. Kluwer.

# The Reaction of Complex Systems to Symmetry Breaking Events: A Mathematical Simulation Model for Neurosciences and Social Sciences

**Pier Luigi Marconi and Maria Petronilla Penna**

**Abstract** The recent Covid-19 pandemic has proposed the issue about how to deal with symmetry break down, by rare events. It is possible that humans have inner skills to deal with such events. The emergence of consciousness may be linked to such resilience and it could support the independence of humans by the determinism of the past, while they are performing the present choices. Awareness may support people to get the best occurrences in the present to reach future goals and to purse their ethical values. However it may be useful, determining the best future goals, to study the behavior of an integrated system when a macrosystem is on failure. In this perspective a math model may be useful to develop. Observations got from neurosciences suggest that wave integration, top down modulation and cross frequency modulation could be all important features to be implemented in such a model. More over the role of the information content and of the catalyst facilitation of the emergent properties, as well as the importance of the integrated complexity, are all very important issues to take into consideration. Wave interactions and information integration are two further properties to be implemented. A multi level system integrating multi component units is proposed.

**Keywords** Brain networks · Cross frequency modulation · Data driven vs. goal driven decision making · Entropy and complexity · Freedom and consciousness · Math modeling of complex systems · Multifrequency complex systems · Sustainable creativity · Symmetry break down · Top-down constructivism in data input

---

P. L. Marconi (✉)
AIRS - Associazione Italiana Ricerca sui Sistemi, Milan, Italy

Artemis Neurosciences, Rome, Italy
e-mail: Pierluigi.marconi@artemisonline.org

M. P. Penna
AIRS - Associazione Italiana Ricerca sui Sistemi, Milan, Italy

Department of Pedagogy, Psychology, Philosophy, Università degli Studi di Cagliari, Cagliari, Italy
e-mail: penna@unica.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_7

## 1 The Black Swan Challenge

Recent pandemic events have made two fundamental phenomena evident.

The first is the emergence of global reactions of biosocial systems to evolutionary processes that modify their complexity.

The second is the effect on the deterministic laws that regulate the intrinsic processes of psychosocial and economic systems determined by acute emergency events.

This effect can reduce or cancel the previously active regularities and linearities, generating singularities that make human predictive systems on failure, experiencing perplexity and disorientation.

These phenomena, referred to as "Black Swan", often occurred in finance, being less pervasive at the psychosocial level, but not less impacting from an economic point of view.

In reality, the appearance of Black Swans in human history is more a rule than an exception, even if their frequency, as the name suggests, is very low, so much so that often it escapes individual life-span observation.

From this it follows that the human mind has evolved in order to develop adequate skills for the management of "Black Swan" conditions, both at the group level and at the individual level.

The essence of the human reaction to the Black Swans consists in the inversion of the cause-effect relationship, which turns from a deterministic condition (where the cause precedes the effect, following a data-driven processing mode) to a finalistic condition (where the cause follows the effect, following a goal-driven one).

This inversion is intrinsic to conscious decision making, where intuition and reason find their integration. Differently it happens when unconscious, instinctive and impulsive reactions occur, where deterministic rules and external conditionings are acting. It is therefore the awareness that allows humans to manage their own freedom from the determinism of the past, aquiring the possibility of reaching the better conditions of adaptation; it is the awareness that support the humans to take into account only the opportunities of the present and the desired values and goals to reach in the future.

Can all these considerations be expressed through a mathematical model, based on systems theory, which allows us to identify which future objectives are more easily pursued in the enhancement of present conditions?

In other words, how can we free ourselves from the conditioning of the past, in the definition of our creative goals, in order to identify them in an equally free and sustainable way?

## 2 The Back Ground of the Model

In this perspective, the possibility of developing a mathematical model based on systems theory could be useful for evaluating which processes "suit" an autonomous intelligent subsystem, if the macro system to which it belongs loses coherence in its behavior.

On this regard, it is necessary to take into account three considerations derived from neuroscience.

- As first, we have the evidence that the brain activity, underlying the mental processes, is fundamentally characterized by modulations of cyclic rhythmic activity, which can be described as wave interactions. Applying mathematical techniques for inverse problem resolution, Steinke and Galán have reverse-engineered network architectures that generate characteristic dynamics of actual brains, including spindles and sharp waves, which appear in the power spectrum as frequency bands superimposed on a non-oscillatory background dominated by low frequencies (Steinke & Galán, 2011). All reconstructed networks display similar topological features and dynamics. By this method the two authors were able to simulate the EEG alteration found in disease. The complexity of the network, quantified as proposed by Tononi, Sporns and Edelman (Sporns et al., 2000; Tononi & Sporns, 2003), appears a good indicator of brain resilience, since virtual brains modeling diseased states display lower complexity than virtual brains modeling normal neural function.
- Secondly, we also have evidence that the activities of higher mental functions, such as consciousness, can generate modifications at the lower levels of the subsystems that underlie them, through a top-down influence. These can occur both in somatic functions modulation and in the perception processing. Specific fronto-temporal cortical regions play a role in the representation and control of adverse conditions, which interact reciprocally with subcortical structures involved in bodily homeostasis and responses to stress (Taylor et al., 2010). Bidirectional autonomic and neuroendocrine pathways transmit information between the central nervous system (CNS) and the periphery, facilitating the expression of affective, autonomic, hormonal, and immune responses (Dum et al., 2019). In perceptual learning, both the encoding and recall of learned information involves a selection of the appropriate inputs that convey information about the stimulus being discriminated (Gilbert & Sigman, 2007). Disruption of this interactions may lead to behavioral disorders, including schizophrenia. Even the brain states are determined by the interactions between multiple cortical areas and by the modulation of multi level feedback connections.
- Finally, we have the evidence that the state of well-being and health tends to be associated with a condition of greater systemic integration and harmonization both at the sub systemic level (in the relationships between body systems and organs) and at the macro systemic level (in the relationships with the surrounding social and natural system). A measurement of the system integration present in the brain may be done relating the phase of EEG band to the amplitude of

a higher frequency band (Cross Frequency Modulation—CfM). The strength of phase-amplitude CfM differs across brain areas in a task-relevant manner, it changes quickly in response to sensory, motor and cognitive events, and it correlates with performance in learning tasks (Canolty & Knight, 2010). Whereas high-frequency brain activity reflects local domains of cortical processing, low-frequency brain rhythms, instead, are dynamically entrained across distributed brain regions by both external sensory input and internal cognitive events. The CfM may be a mechanism to transfer information from large-scale brain networks operating at behavioral timescales to the fast, local cortical processing required for effective computation and synaptic modification, thus integrating functional systems across multiple spatiotemporal scales (Canolty & Knight, 2010). Alterations of CfM were observed to correlate with pathology, in schizophrenia (Allen et al., 2011) and in people maltreated in infancy and adolescence (Marconi et al., 2018).

From these considerations emerges the interest to focus the study of the behavior of a complex system, composed of a set of subsystems and macro systems, each characterized by specific wave functions. In particular, the integrations of larger systems (in terms of spatial coordinates) can be modeled by wave functions with a lower frequency but higher propagation speed, while the integrations of smaller systems by waves with a higher frequency but also lower propagation speed. This corresponds to what we observe at the cerebral level in the integration processes. Von Stein and Sarnthein have found that local synchronization during visual processing evolved in the gamma frequency range, while synchronization between neighboring temporal and parietal cortex during multimodal semantic processing evolved in a lower, the beta1 (12–18 Hz) frequency range, and long range fronto-parietal interactions during working memory retention and mental imagery evolved in the theta and alpha (4–8, 8–12 Hz) frequency range (von Stein & Sarnthein, 2000). The authors suggested that a relationship could exist between the extent of functional integration and the frequency of synchronization. In particular, long-range interactions in the alpha and theta ranges seem specifically involved in internal mental processing and top-down processing.

These electrical activities integrate single cortical areas in more complex systems, called brain networks (Canolty & Knight, 2010), characterized by the same macro function (DMN—default mode network, CEN—central executive network, SN—salience network). In a paper published in 2018, Marconi, Penna e Pessa matched real data got from maltreated people or control subject with a math model of brain network integration. The external inputs influence both CEN (involved in controlling reactions to external inputs) and DMN (involved in processing of internal stimuli and memories) in an indirect way, mediated by the relationships between them and the SN (involved in valence attribution of inputs and in the activation of the related brain networks). This is concordant with the hypothesis that, even in presence of dysfunctions regarding the latter two networks, the brain could compensate for their faulting performances owing to the presence of collateral contributions coming from the eventual inputs (Marconi et al., 2019).

And this interpretation seems to be confirmed by the available experimental and clinical data.

Particular (and extreme) values attributed to the numerical coefficients present in the math formulae of the model were used to simulate two possible situations.

1. If both CEN (i.e. the conscious relationship with the external world) and DMN (i.e. the relationship with memories and the inner world) networks are characterized by extreme deficits of their activity, the brain activity is mainly ruled by Salience network, stimulated by inner emotional states and /or influenced by inputs contribution: a situation characterizing many psychiatric and neurological disorders.
2. On the other hand, if both SN and DMN networks are characterized by extreme deficits of their activity, the only non-zero value of equilibrium amplitudes is the one of SN, if still a brain function could be identified.

Clinical and simulation data were confirming the hypothesis that for people with personal history of child maltreatment the CEN is activated to force a top-down effect which switch the valence attribution from the SN/DMN network interaction to a more aware processing SN/CEN network interaction, the first one making reference to previous memories and the second one to formal properties of inputs.

## 3 Preliminary Considerations for Modeling

It is very important to keep in mind the difference between basic physical phenomenon and information management process.

In fact, the architectural and functioning of a resilient system is fundamentally based on three aspects:

1. The presence of a relationship between components with the maximum probability of persistence over time
2. The emergence of a property (acting as an "organizational field"), which increases, as a catalyst, the speed of all the subsystems to converge towards that high probable functional organization (with top-down influence),
3. The complexity of a system, intended as information content ("integrated information").

The importance of information contained and managed by a system has already been highlighted by Giulio Tononi, in his theoretical work that link the emergence of consciousness to the quantity of integrated information, i.e. the greater quantity of information contained in a system that can be derived from process of integration of its components (Tononi, 2005).

Probability of persistence, facilitated states and complexity of a system are all concepts that can be linked to the statistical perspective of the concept of entropy.

In spite of the fact, that entropy was originally defined by German physicist Rudolph Clausius as the quotient of an amount of heat, in the dissipative use of energy during a transformation, a new definition was developed by Ludwig Boltzmann in the 1870s that proposed a statistical definition of entropy by analyzing the statistical behavior of the microscopic system components. In statistical mechanics the interpretation of entropy is the measure of uncertainty about a system, which still exists after its observable macroscopic properties, such as temperature, pressure and volume, have been taken into account. In such a perspective, entropy measures the degree to which the probability of the system is spread out over different possible microstates, given a set of macroscopic variables. Finally, when viewed in terms of information theory, the entropy state function is the amount of information that is needed to fully specify the microstate of the system. This is lacking in the macroscopic description.

Thus the phenomenon of entropy, which characterizes the second law of thermodynamics and which seems to condemn us to an inevitable death, can actually be seen, instead, as a process of increasing complexity of a system, a process that correlates each other time, integration, persistence and information content increase of a system.

The importance of dealing with information carried by waves can also be derived from the paradox mathematically observable in the function that represent the interference between waves of different phase, speed and frequency. It can be easily seen how the information carried by the interference effect can be transferred more quickly than the maximum speed of the original waves, which highlights how the information transported by two waves can travel at a speed even faster than light. The phenomenon of entanglement is also fundamentally based on the transmission of information. The term was introduced by Erwin Schrödinger (1935). Quantum entanglement, or quantum correlation, is a quantum phenomenon, not reducible to classical mechanics, for which under certain conditions two or more physical systems represent subsystems of a larger system whose quantum state cannot be described individually, but only as a superposition of more states. From this it follows that the measurement of an observable subsystem simultaneously determines the value for the others as well. Since it is possible from an experimental point of view that such systems (subsystems) are spatially separated, entanglement implies in a counterintuitive way the presence of distant correlations, theoretically without any limit, between their physical quantities, determining the non-local character of the theory (Moreau et al., 2019).

## 4 The Proposed Model

All these considerations make very relevant to consider information content, complexitity, facilitation (probability), wave functions and interference, simulating a resilient multilayer system, integrated with larger systems. Based on these considerations, it becomes possible to represent the complexity of an integrated system as a multilevel set of systems, described by waves, as information carriers.

- The proposed model, has five layers of integrated systems.
- The layers are integrated by wave interference functions and cross frequency modulations, as well as they are activated by the resultant high frequency activity of the lower layers and modulated by the low frequency activity of higher levels.
- The fourth and fifth layers represent the external environment, acting by an entrainment like effect. Entrainment is the effect that external rhythms can have on neural activity, leading a coherent internal rhythm with the external inputs. Such a coherence can works as a CfM in the integration of individuals in a global natural system.
- Interruptions or disruptions in the rhythm of the fourth layer simulate strong external events as the Black Swan, as well as positive or negative occurrences.
- The persistence of the rhythm in the fifith layer simulate the persistence of symmetries at higher systemic levels, to which the “four layer environment” belongs.

This proposed model can allow us to evaluate theoretically the necessary characteristics of a subsystem, to guarantee its possibility of finding new ways of integration, when the meso system in which it is integrated finds a condition of singularity and symmetry breaking.

In particular, this model could highlight which conditions increase the probability of persistence (life) over time after a Black Shown, 1) or by bringing out new meso systemic organizations with a bottom-up modality, 2) or by quickly responding to a meso systemic reorganization induced by the higher macro system with a top down mechanism, 3) or both.

The highlighting of the theoretical issues and constraints of this optimal reaction will allow us to better identify, also on the rational level, as well as on the instinctive reaction and/or behaviors derived from ancient cultural traditions, what is the best goal and the best system of values to be used in the inversion of the cause-effect relationship that characterizes the conscious response to events of the “Black Swan” type.

## References

Allen, E. A., Liu, J., Kiehl, K. A., Gelernter, J., Pearlson, G. D., Perrone-Bizzozero, N. I., & Calhoun, V. D. (2011). Components of cross-frequency modulation in health and disease. *Frontiers in Systems Neuroscience, 5*, 59. https://doi.org/10.3389/fnsys.2011.00059. PMID: 21808609; PMCID: PMC3139214.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences, 14*(11), 506–515. https://doi.org/10.1016/j.tics.2010.09.001

Dum, R. P., Levinthal, D. J., & Strick, P. L. (2019). The mind-body problem: Circuits that link the cerebral cortex to the adrenal medulla. *Proceedings of the National Academy of Sciences of the United States of America, 116*(52), 26321–26328. https://doi.org/10.1073/pnas.1902297116. Epub ahead of print. PMID: 31871146; PMCID: PMC6936592.

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron, 54*(5), 677–696. https://doi.org/10.1016/j.neuron.2007.05.019. PMID: 17553419.

Marconi, P. L., Penna, M. P., & Pessa, E. (2019). The psychopathological process as a system of dysfunction and systemic compensation with top-down modulation. In G. Minati, M. Abram, & E. Pessa (Eds.), *Systemics of incompleteness and quasi-systems. Contemporary systems thinking*. Springer. https://doi.org/10.1007/978-3-030-15277-2_14

Marconi, P. L., Pessa, E., & Penna, M. P. (2018). Comparing different methods of measurement in psychiatry and clinical psychology: The diagnostic power of the new neurophysiology techniques. In *IEEE International Symposium on Medical Measurements and Applications (MeMeA),* Rome*,* 2018 (pp. 1–6.) https://doi.org/10.1109/MeMeA.2018.8438667.

Moreau, P. A., Toninelli, E., Gregory, T., Aspden, R. S., Morris, P. A., & Padgett, M. J. (2019). Imaging bell-type nonlocal behavior. *Science Advances, 5*, eaaw2563. https://doi.org/10.1126/sciadv.aaw2563

Schrödinger, E. (1935). Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society, 31*(4), 555–563. https://doi.org/10.1017/S0305004100013554

Sporns, O., Tononi, G., & Edelman, G. M. (2000). Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex, 10*, 127–141.

Steinke, G. K., & Galán, R. F. (2011). Brain rhythms reveal a hierarchical network organization. *PLOS Computational Biology, 7*(10), e1002207. https://doi.org/10.1371/journal.pcbi.1002207

Taylor, A. G., Goehler, L. E., Galper, D. I., Innes, K. E., & Bourguignon, C. (2010). Top-down and bottom-up mechanisms in mind-body medicine: Development of an integrative framework for psychophysiological research. *Explore (New York, N.Y.), 6*(1), 29–41. https://doi.org/10.1016/j.explore.2009.10.004

Tononi, G. (2005). Consciousness, information integration, and the brain. *Progress in Brain Research, 150*, 109–126. https://doi.org/10.1016/S0079-6123(05)50009-8. PMID: 16186019.

Tononi, G., & Sporns, O. (2003). Measuring information integration. *BMC Neuroscience, 4*, 31. https://doi.org/10.1186/1471-2202-4-31. PMID: 14641936; PMCID: PMC331407. Complexity toolbox. http://www.indiana.edu/~cortex/complexity.html.

von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology, 38*(3), 301–313. https://doi.org/10.1016/s0167-8760(00)00172-0. PMID: 11102669.

# What Would I Propose to Eliano to Continue Our Collaboration: Validity Regimes, Edges, Interfaces, and Waves of Complexity

**Gianfranco Minati**

**Abstract** The purpose of this contribution is to introduce and elaborate some research topics that Eliano and I had in progress. They are the concepts of validity regimes of symbolic, sub-symbolic, and quasi-rules, edges, transients, and waves of complexity within the dynamics of complex systems. Such concepts are introduced as possible approaches to model the irreducible, analytically non-zippable, and theoretical incompleteness and quasi-ness of complexity. We consider the continuous trade-off between coherence—incoherence, incompleteness—quasi-ness, levels of emergence, the infiniteness of betweenness, and collapsing mechanisms of complexity. The complexity of emergence is considered as coherent dynamics of edges, interfaces of validity regimes, and their transients. We mention an elementary example dealing with analytical representations with ordinary differential equations—completeness context—rather than with sub-symbolic or network cases. They are elaborated upon here as they had to be colloquially and informally presented to Eliano for his consideration and further reflection.

**Keywords** Betweenness · Coherence · Edges · Incompleteness · Interfaces · Levels of emergence · Quasi-ness · Transients · Validity regimes · Waves of complexity

## 1 Introduction

The AIRS[1] experience and my dedication to research after a career as an executive in a large Italian financial and industrial group started with my meeting Eliano

---

[1] Associazione Italiana per la Ricerca sui Sistemi, in English Italian Systems Society http://www.airs.it/AIRS/indexEN.htm.

G. Minati (✉)
Italian Systems Society, Milan, Italy
e-mail: gianfranco.minati@AIRS.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_8

around 25 years ago. This part of my life started some years after losing my wife and my multiple sclerosis diagnosis. I opened with Eliano the doors of this new life, and from then on, we always traveled together. Eliano had different lives, one in support of the others. Theoretical physicist, cognitive scientist, systems scientist, and professional mountain climber. Interestingly my increasing disability and reduced ability to walk, in some ways combined with his mountaineering attitude. We both had climbed different mountains until now.

We dedicated ourselves to various issues, always of a systemic nature, sharing our findings in articles, books, and conferences, covering various topics, including but not limited to; collective beings, constructivism and cognitive operators, dynamic use of models, emergence, incompleteness, logical openness, meta-structures,[2] multiple-systems, post-Bertalanffy systemics, quasi-ness, and quasi-systems. Pursuing these ideas to some perceived frontier that I would like to, although certainly inadequately, consider here in honor of Eliano, such as validity regimes, their edges, interfaces, and transience; dynamics as transience among validity regimes, and structural dynamics.

We were exploring *respectful* approaches to complex systems and not increasingly sophisticated ways of prescribing them properties and behaviors, but rather ways to orient and induce, in essence, *proposals* to be elaborated by the complex systems. We realized the enormous inadequacy of the theoretical tools available born based on completeness in the decision-making *presumptuous* scheme.

I outline, to my knowledge level, a related ecosystem of concepts and approaches. It is a sort of task inherited from Eliano by the AIRS and myself. I will try to present concepts and approaches such as stones that we meet on a climb. Knowledge accrued, not in the invasive excavation of a mine, but as respectfully climbing, through frequenting materiality and ideas.

It would be an impressionist, theoretical incomplete, logically open scenario phenomenologically driven such as climbing unexpected mountain paths.

I will try to do it considering having him beside even if he is climbing other unimaginable mountains.

## 1.1 About the Chapter

This book is dedicated to Eliano, to what we have done, particularly its recognizable logical openness. I write this contribution as a fact, as far as I am concerned, of dutiful specification of the theoretical unfinishable openness, unfinishable sequences of non-equivalent models, as a matter of theoretical incompleteness (Minati, 2016a).

This contribution is a kind of tribute to our conceptual focus, from the content in our recent publications on multiplicity, quasi-ness, and incompleteness where the

---

[2] Structures among clusters (Minati & Licata, 2012, 2013, 2015; Minati et al., 2013; Minati & Pessa, 2018, pp. 102–129).

phenomenological freedom of becoming (different from randomness) is intended to be. It is a matter of logical openness when *modeling* may be only partial rather than *complete*, and the strategy of being able to exhaust is ineffective rather than just wrong. However, as a conceptual framework, the non-necessary homogeneous, complete explicability of Nature, must first be understood. The coexistence and overlapping of different, non-equivalent approaches are assumed to correspond to its constitutive multiplicity, as in, climbing several different routes is possible. Such considerations are reminiscent of the previously introduced concepts of logical openness (Minati et al., 1998; Minati & Pessa, 2006, pp. 64–75) and the DYnamic uSAge of Models (DYSAM) as in (Minati & Pessa, 2006, pp. 64–75; 2018, pp. 201–204).

The complex, inhomogeneous, and incomplete explicability of Nature, is coherent with Eliano's and my understanding. This is contrasted by the objectivistic one, self-considered as *external* and presumptuously independent from it (Minati, 2019a), as it was not a natural fact of self-understanding.

*I mention how most of the related co-authored articles and books have me as the first author. Beyond the alphabetical order, Eliano's natural tendency towards modesty and confidentiality, undeservedly, these credits are attributed to me*.

The purpose of this contribution is to present unfinished, conjectured approaches based on previously introduced theoretical frameworks. To better understand the irreducible, analytically non-zippable, theoretical incompleteness (Minati, 2016a; Longo, 2011) and quasi-ness of the processes of emergence in complex systems. We mention how this conceptual framework is entirely different when considering other approaches, where incompleteness is not theoretical but *conceptually completable*, phenomenological, e.g., due to measurements, rather than necessary for becoming, for emergence. For example, the so-called *grey systems theory* is characterized by incompleteness in information as one of the characteristics of uncertain systems, such as when incomplete system information relates to the fact that the information about the elements, the structure, the boundary, and the system's behaviors is incomplete (Javanmardi et al., 2020; Liu & Yang, 2012). Another well-known case relates to fuzzy systems whose elements have membership degrees within the continuous interval between 0 and 1 (Zadeh et al., 1996).

I dedicate this chapter to topics such as validity regimes of symbolic, subsymbolic, and quasi-rules (when irregularly and in different ways apply); edges and transients as considered in our last book (Minati & Pessa, 2018). They are elaborated upon here as they had to be colloquially and informally presented to Eliano for his consideration and further reflection. Instead of Eliano, they are presented here to the systemic scientific community for further developments, simulations, and applications, but at least in honor of Eliano. The complexity of emergence is considered here as coherent dynamics of edges, interfaces of validity regimes, and their transients.

Domains wherein some properties are exclusive, unique, or almost prevalently dominant are intended here as *validity regimes* (Minati & Pessa, 2018, p. 88, pp. 127–130, pp. 265–266). Examples are domains of entities collectively interacting through specific or combinations of interaction mechanisms. Their interacting makes their behavior to assume properties, for instance, chaotic, ergodic, correlated, networked, polarized, and scale-invariant.

Elementary cases of validity regimes' edges occur when edges are constituted by a single or few analytically well-composed predominant validity regimes where there is adjacency, any partial combination, with other validity regimes of which in turn constitutes the edge. Thus, edges can only be multiple unless they are edges towards the rarefaction and the validity regime's disappearance. The internal predominance is only parametrically variable, allowing compatibility and may also have the role of initial conditions for processes of transience converging towards a new validity regime as in metastability (Slowik, 2012). In cases of turbulent regimes when predominance is unstable and insignificant in percentage, edges are probably undetectable and irrelevant.

A different case of interest here consists of considering the generic complex behavior as having different levels and types of coherence and moving edges of multiple validity regimes when considering their topological, predominant, and mixing aspects establishing multiple, well- or quasi-ruled domains.

In conceptual correspondence with propagation phenomena in physics and to the so-called *moving boundary problems*, describing, for example, the solid-liquid interfacing during phase transitions, we consider edges of validity regimes of interest because representing the internal, structural dynamics of complex systems and their multiple natures.

Classic examples of edges are the frontiers of generic and volcanic eruptions, studying how they spread, in case mix, in the environment, and seismic waves. In complex systems, e.g., in flocks and swarms composing entities, variable in number and with different topological properties, are considered constituting edges when the predominant interaction mechanisms significantly change, e.g., increasing or decreasing in intensity, replaced or combined with different ones, such as at the boundaries, at the center, and due to environmental interferences. When there are no replacements and combinations, the constitutive interactions may have reduced validity and predominance, cease of validity temporarily, be in transforming (e.g., adaptive) phase, or degenerating towards disaggregation.

*Dealing with collective phenomena, composing entities' phenomenological interaction mechanisms are usually neglected or, rather they are inferred, re-engineered or replaced by considering simulation rules (*Minati & Pessa, 2006*, pp. 104–105). Phenomenological interaction mechanisms are otherwise considered represented by the validity of acquired, generalized global properties, such as having chaotic, correlational, scale-invariant, and ergodic natures. We consider here validity regimes and their edges where such multiple interaction mechanisms and properties of different natures variably combine and coexist, have variable predominance, diffusion, and overlapping, allowing the conjecture of waves of complexity as dynamics of complexity. We may have validity regimes specified by interaction mechanisms together with possible modalities of change, e.g., ergodically or networked, acquired by configurations of interacting entities; modalities of change applied by configurations of entities; and interaction* *mechanisms with no specific modalities of change applied by configurations of interacting entities.*

*We consider the dynamics of complex, multiple, quasi-systems having multiple processes of emergence as dynamics of edges and transience between validity regimes*. A particular case of this view is the dynamics of radical, multiple emergences occurring between multiple processes of the phase transition-like processes. As a typical case, we mention ecosystems of multiple, simultaneous, interacting transients among validity regimes inferring or representing their phenomenology.

**In conceptual correspondence with physics, where the discovery of electromagnetism introduced the opportunity to consider the fields as primary entities rather than objects, we try here to consider validity regimes and their properties as primary entities of complex systems rather than their phenomenologically interacting materiality** (Minati, 2019a).

We mention in the following the contribution of the different sections.

In Sect. 2, we introduce the concept of a validity regime, initially intended as the domain of admissible subsequent states, and then wherein multiple rules of interaction (inferred interaction mechanisms) or specific acquired properties of changing prevail. Cases of significant validity regimes (Minati & Pessa, 2018, pp. 127–130, pp. 265–266) are the inferred interaction mechanisms, e.g., simulations, and the chaotic, dissipative, networked, and scale-invariant ones. They may be combined and suitably parametrized.

In Sect. 3, we consider quasi-ness to deal with analytically intractable ecosystems of properties, allowing the emergence of properties and coherence. The concept of quasi-ness pertains to instabilities of properties, non-regular alternations of degeneration and recovery of properties occurring when a system is not always a system, not only a system, not always the same system, and considering levels of similarities and equivalences (Minati, 2018, 2019a, b, c; Minati & Pessa, 2018; Minati et al., 2019). We present an elementary case based on ordinary differential equations as an example of quasi-systems and their related validity regime.

In Sect. 4, we consider the edges of validity regimes in conceptual correspondence with the propagation phenomena known to physics. In systems science, the edges are typically multiple systems[3] subjected to multiple interaction regimes that acquire emergent stabilities and roles, e.g., topological and energetic. We consider the interfacing role of their edges, transience as on-going betweenness and interfaces, and in Sect. 5, the dynamics of complex systems intended as transience among validity regimes.

In Sect. 6, we introduce issues for further research.

In the conclusions, we present some possible consequent research approaches. Specifying and reassuming the discussion that I wish I could have with Eliano.

---

[3] Multiple Systems are given by the multiple roles of their constituting interacting components; by interchangeability among components which take on the same roles at different times and different roles at the same time; by the occurrence of multiple partial mediated flows of information through components; and by multiple interactions such as multiply linked nodes in Networks (Minati & Pessa, 2018, pp. 42–45; pp. 166–170). A classic example is that of ecosystems.

## 2 Validity Regimes

The concept of the domain considered here, and then constituting the one of validity regime when related to predominant interaction mechanisms and potential acquired properties of the resulting change in interacting entities, is different from the one used in some disciplines such as in mathematics considering the domain of a function, in Information Technology when dealing with web domains and magnetic domains in physics intended as regions of a magnetic material in which the magnetization is uniform.

*The domain (*Minati, 2019a*) of an entity is initially considered, given by the admissible subsequent interconnected values that variables representing the degrees of freedom can assume in the following temporal configurations. In this case, the basic concept of the domain considered relates to single entities, their possible compositions such as configurations and systems when their admissible domains may have the property to keep behavioral features, e.g., adaptability, consistency, and functionalities*.

A single entity or configurations of entities and their previous behaviors are assumed to 'generate' their own admissible domains and use them during the dynamical evolution. However, the domain should not be reduced to possible deterministic changes, occurring when the dynamics of the evolutionary behaviors are intended to be *completely* represented in a context of *logical closedness*. Conversely, complex systems evolve in a context of *logical openness* when such completeness is not possible, see Minati et al. (1996, 1998), Minati (2016a), and Minati and Pessa (2018) and processes of emergence as the acquisition of new (irreducible to the previous ones) configurations, occur in phase-transition ways.

The domain is then variable over time since entities and their configurations at time $t_n$ 'generate' their domain valid for $t_{n+1}$ allowing options even previously inadmissible. The following options available may be non-linearly successive, non-equivalent such as succeeding configurations of self-organization and emergence processes. However, such non-complete domains are quasi-compatible when acquired, and non-equivalent properties preserve and resume variable and coexisting coherence levels. In case their predominance becomes convergent and unique for second-order phase transitions.

On the other hand, we may consider *potential* domains without generating reference entities but having characteristics corresponding to multiple, dynamic *environmental* constrained behavioral rules. This is in conceptual correspondence with considering environmental properties such as for ecosystems when "…the environment *pervades* the elements which produce, in their turn, an active environment. This environment, if we can still call it such, is active and not an amorphous, abstract space hosting processes. It is interesting to consider eventual conceptual correspondences with the quantum vacuum pervading everything." (Minati & Pessa, 2018, p. 13).

The concept of a domain constitutes the one of validity regime when the reference is not anymore to the admissible interconnected changes only, values that variables representing the system may acquire, but to modalities and properties of changing for entities.

*Validity regimes are intended as domains that constitute interaction mechanisms, possible modalities, and properties of changing according to which configurations of entities occur.*

A simple example, is given by a validity regime where specific interaction mechanisms[4] apply and the evolution of dynamic configurations evolve with properties, for instance, ergodically or in correlation. They are properties of *any* or, more realistically, *predominant* changes or changing due to applying the rules such as specific interaction mechanisms. Cases of such properties include the coexistence, variable predominance, and combinations of chaotic, ergodic, correlated, networked, polarized, and scale-invariant regimes.

The domain of rules, that is, the validity regime, of an ecosystem is intended as constituted by multitudes of properties of behavioral rules having compound, fuzzy, missed origins; reproduced, iterated, combined, or modified for any reason.

*We stress that we consider the term 'rules' in extended ways, such as architectural properties, e.g., levels and weights of Neural Networks; clusterizations, networking parameters and properties; and combinations of multiple analytical representations.*

This is in conceptual correspondence with Quantum Field Theory (QFT), where the void is not emptiness, absence of everything, but a pervasive, unavoidable source of properties such as entanglement (Gühne & Toth, 2009). Furthermore, in the quantum vacuum lacking any particles, there are fluctuating electromagnetic fields, fluctuating about an expectation value of zero. The quantum vacuum is intended to precede matter, and in such a way, it also must precede space and time (Preparata, 2002). Quantum properties, such as entanglement, may be considered to establish validity regimes.

We conclude this section summarizing, see Table 1, that validity regimes are intended as domains where inferred interaction mechanisms and properties of multiple, combined rules of changing apply.

## *2.1 Validity Regimes of Populations of Interacting Entities*

Considering a population of interacting entities (for a review, see Vicsek & Zafeiris, 2012), a validity regime is deemed to be established when the interaction

---

[4] We mention, from Reynolds (1987), the very popular interaction mechanism used for flock simulations constituted of: separation rules (individual boid must control their motion in order to avoid collisions and the crowding of locally adjacent components); alignment rules (individual boid must control their motion so as to point towards the average motion direction of locally adjacent components); and cohesion rules (individual boid must control their motion so as to point towards the average position of locally adjacent components) implemented in several simulators such at http://sourceforge.net/projects/msp3dfbsimulator/?source=directory.

**Table 1** From domains to validity regimes

| |
|---|
| *Domains of single entities*: admissible states assumable over time. Domains changes over time accordingly. |
| *Domains of configurations of interconnected entities*: admissible interconnected changes assumable over time. Domains changes over time accordingly. |
| *Domains of collective entities*: validity regime as the occurring of inferred interaction mechanisms possibly together with modalities and properties of change for configurations of entities such as having chaotic, correlational, scale invariant, and ergodic natures. |

mechanisms are supposed, inferred to occur through a single or composition of rules, stable or variable in any way, and when occurring according to properties and modalities such as correlation, ergodicity, and networking.

However, the concept of validity regime is spatially generic; it may apply to disconnected spaces and sets of fibers depending on the representation level. Furthermore, we may consider spaces of signals and economic transactions.

Validity regimes of rules and properties (since now only 'validity regimes' for short) establish the interactional environment, characterizing the occupied space and to which each new entity involved must adapt, similarly to enter the field's validity domain and that, even if partially, the entering entity must fulfill. However, as we will see, validity regimes may be analytically completely well, multiple, or quasi-defined when rules and properties apply irregularly, partially, and inhomogeneously.

**The generic introductory concept of validity regime elementary starts by considering a domain where a single or few interaction mechanisms are supposed to apply homogeneously**. Such validity regimes may be ideal and ***reversely engineered***, ***interpolated,*** **or** ***abductively***[5] ***inferred*****—a typical example is the classical Reynolds model** (Reynolds, 1987) **introduced for graphic simulations (see Footnote** 4)**—and in dependence of experimental properties detected** (Herbert-Read et al., 2011).

A very well-known ideal, simplified validity regime introduced by Ludwig von Bertalanffy (1901–1972) is considered as the idealized prototype of the very concept of a system. He considered a system as ideally specified by suitable state variables $Q_1$, $Q_2$, …, $Q_n$ whose instantaneous values specify the system's state (von Bertalanffy, 1968, p. 56). Examples of $Q_i$ are the temperature and pressure (representing, for instance, thermodynamic systems like steam engines) and functional, depending on the nature of the system. Also, the classical *general* analytical representation of collective interactions establishing (in case, self-organized or emergent)[6] system

---

[5] In logics abduction is the inference, the process of forming an explanatory hypothesis.

[6] Self-organization intended as sequence of properties acquired in a phase-transition-like manner having regularities and synchronizations, e.g., whirlpools and tornados. Emergence when the sequence of properties acquired in a phase-transition-like manner is *coherent*, e.g., flocks and swarms (Minati & Pessa, 2018, pp. 65–86; pp. 255–260; Minati, 2019a). In this view regularities and repetitiveness of self-organization are particular cases of the *coherence* of emergence (Minati, 2016a).

characterized by suitable state variables $Q_1, Q_2, \ldots, Q_n$ is given by their time evolution ruled by a system of ordinary differential equations, such as:

$$\begin{cases} dQ_1/dt = f_1\,(Q_1, Q_2, \ldots Q_n) \\ dQ_2/dt = f_2\,(Q_1, Q_2, \ldots Q_n) \\ \ldots \\ dQ_n/dt = f_n\,(Q_1, Q_2, \ldots Q_n) \end{cases} \tag{1}$$

In the case of collective systems, examples of $Q_i$ are density, levels of internal correlation, volume, shape, and their ways to change (representing, for instance, collective behaviors of flocks, swarms, and markets) depending on the nature of the system. This system of ordinary differential equations specifies how a change in the value of a given state variable $Q_n$, is related to all other state variable changes through $f_n$, representing their *collective simultaneous interdependence*.

The validity regime coincides with the interaction mechanism when $Q_n$ specifies features of a single entity's behaviors, such as in flocking the distances between boids, directions, and speeds of their motion, as in Footnote 4. We may consider $[Q_n]$ as a vector representation of the population dynamics. *In this case, the inferred interaction mechanism is completely well defined, and the correspondent validity regime as well.* This is an idealistic, unrealistic, yet very illustrative approach.

Well-known examples of equations stating *predominant,* inferred interaction mechanisms and their properties, suitable to identify corresponding validity regimes and compatible with levels of complexity, are those of chaotic regimes, specified by the equations of chaotic behavior (Lorenz equations); networking regimes specified by properties such as the *local clustering* and *small-world properties*; scale-invariant regimes; correlational regimes where changes are predominantly correlated, and regimes having significant levels of ergodicity.

The previous are all examples of well-defined validity regimes which, however, could be composed in multiple and irregular ways. When the interaction mechanisms and the properties of the complex, emergent behavior are completely represented and well-defined, it is a matter of fortunate cases often considered, however, as *essential ideal generative mechanisms of which the phenomenological ones are only to be understood as inaccurate and approximate variations.*

A simple opposite case takes place when interaction occurs unruled, in apparent disordered ways but acquiring some statistical properties, for instance, induced by constraints. An example is given by the disordered interacting of gas molecules having thermodynamic statistical properties. In this case, the validity regime is just a list of on-going rules with no unifying properties.

**However, instead of well-defined and stable ruling and properties, we may infer the occurrence of their quasi-ness, i.e., irregular occurrence, superimposition, unstable coherence, different durations, and starting time.** In fortunate cases, it is a matter of continuous, oscillatory reductions, increasing in coherence, loss, and recovery, not realistically constant, as it is instead predominant in self-organization cases (see Footnote 5). An example is given by multiple interrelated, networked environments such as ecosystems, where external and internally-generated inputs continuously apply.

In such cases and radical emergent complex systems, the representation of the phenomenological interactional mechanisms is not analytical but adaptive, ongoing, like natural computation (Brabazon et al., 2015).

An interesting case takes place when the comply with the rules of the validity regime by whatever entering entities makes in their turn a population of interacting entities to acquire properties, for instance, induce a phase transition and break metastability states (Slowik, 2012).

In sum, we may consider:

- Validity regimes as domains of possible available collective (e.g., density) and global states (e.g., shape);
- Validity regimes as interaction mechanisms and properties of the evolutionary paths of the behavior of the elements. Validity regimes:
  - Idealized,
  - Inferred from experience,
  - Abductively, reversely engineered or interpolated, explanatory of the properties acquired,

  Rarely experimentally detected.

## 3 Quasi-ness of Validity Regimes

The concept of *validity regime* we considered as suitable to deal with cases of radical emergence related to the occurrence of unique multiple coherences, when numerous, superimposed processes of emergence occur and nor rules neither properties are completely, definitively defined or well separable and no significant predominance characterizes the dynamics.

Detected properties may be supposed due to and according to applying some composed interaction mechanisms establishing the validity regime.

The following considerations may be worthy of note:

- Validity regimes characterized by a significant temporal and spatial predominance of some rulings and some related global properties; if not, we may speak of *weak* (Minati, 2016b) *validity regimes*.
- Validity regimes include the rulings of quasi-systems and multiple systems (see Footnote 3).
- Validity regimes are trivial when rulings are entirely well-defined and fixed.
- Being a multiple system does not imply its quasi-ness. Anyway, we introduce in the following some specifications about quasi-ness of validity regimes. Such specifications are introductory to deal then with *interfacing edges*, waves of edges, and their dynamics as transience among validity regimes considered in Sects. 4 and 5 and allowing to consider the concept of *waves of complexity*.

## 3.1 An Elementary Analytical Example of Quasi-ruled Validity Regimes

The interest in quasi-ruled validity regimes is given by their compatibility, probably necessity even not sufficiency to induce, to establish emergence processes. The sufficiency may require the introduction of suitable constraints. *The multiple coherences of emergent systems are given and require multiple equivalences; their trade-off; balance and unbalance; loss and recovery of property not reducible to a single, iterated, specific coherence.*

We present a generic elementary case however suitable to represent the general idea of multiple roles, quasi-ness, and reduction of the number of degrees of freedom.

We will take into count the original approach considered in Sect. 2.1. Consider a population of interacting elements for which suitable state variables $Q_1, Q_2, \ldots, Q_n$ instantaneous values specify the state of the complex collective system.

In simple cases, the system of ordinary differential equations (Eq. 1) fully represents the population dynamics when behaving as a *perfect*, and as such weakly complex, structure.

Quasi-systems relate to different possible variants when, for instance, the interaction mechanism non-homogeneously applies; it applies in different ways, with different parameters; different compatible interaction mechanisms irregularly, but with predominant coherence, applies.

Generic versions of such variants may be considered.

As a first elementary case, however useful to give the idea of the approach, we may consider two initially separated, i.e., not interacting and occupying spatially separated areas, populations of the same kind of composing entities, such as swarms of same insects and markets of same goods in different places or time.

We may start by considering, for example, their two well-defined system regimes, that is, two systems such as:

$$\begin{cases} dQ_1/dt = f_1\,(Q_1, Q_2, \ldots Q_n) \\ dQ_2/dt = f_2\,(Q_1, Q_2, \ldots Q_n) \\ \ldots \\ dQ_n/dt = f_n\,(Q_1, Q_2, \ldots Q_n) \end{cases} \tag{2}$$

and

$$\begin{cases} dQ_1/dt = g_1\,(Q_1, Q_2, \ldots Q_n) \\ dQ_2/dt = g_2\,(Q_1, Q_2, \ldots Q_n) \\ \ldots \\ dQ_n/dt = g_n\,(Q_1, Q_2, \ldots Q_n) \end{cases} \tag{3}$$

The two systems (Eq. 2) and (Eq. 3) consider the same state variables $Q_1, Q_2, \ldots, Q_n$. and differ for their structural interdependence, i.e., $f_n$ and $g_n$, such as mechanical and electromagnetic, or interaction mechanisms, e.g., considering spatial positions, distances and speeds of elements.

This simplified example may consider several variants, conceptually starting from cases when $f_n$ in Eq. (2) and $g_n$ in Eq. (3) representing the single interaction mechanisms of systems' elements, apply in different ways when, for instance, they play roles in the *other* interaction mechanism. For instance, they may irregularly replace, due to overlap or preponderance, the other one or combine as in Eqs. (4) and (5):

$$\begin{cases} dQ_1/dt = f_1\,(Q_1, Q_2, \ldots Q_n) \\ dQ_2/dt = g_2\,(Q_1, Q_2, \ldots Q_n) \\ \ldots \\ dQ_n/dt = f_n\,(Q_1, Q_2, \ldots Q_n) \end{cases} \tag{4}$$

$$\begin{cases} dQ_1/dt = g_1\,(Q_1, Q_2, \ldots Q_n) \\ dQ_2/dt = f_2\,(Q_1, Q_2, \ldots Q_n) \\ \ldots \\ dQ_n/dt = g_n\,(Q_1, Q_2, \ldots Q_n) \end{cases} \tag{5}$$

This may occur for any reason, such as environmental influences, energetic variations, and interferences[7] between the two systems that find themselves overlapping, crossing, and sharing the same environment and roles. Conceptually, it is the case in physics when multiple fields variably apply and superimpose.

A further generalization can be introduced, considering the following systems, $S^1$, $S^2$, and $S^3$ and subsequent possible $S^n$ as in Eqs. (6), (7), and (8). **However, different associations between the systems and their state variables represent elementary multiplicity and quasi-ness of the validity regime under consideration specified in** Eq. (9) (Minati & Pessa, 2006, pp. 123–128).

$$S^1 : \begin{cases} dQ_1/dt = f_1\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_2/dt = f_2\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_3/dt = f_3\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_5/dt = f_5\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_6/dt = f_6\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_7/dt = f_7\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \end{cases} \tag{6}$$

---

[7] While the interaction changes the behavior of the individual entities, the interference changes the influences of the interaction or the interaction itself.

$$S^2 : \begin{cases} dQ_4/dt = f_4\,(Q_4, Q_5, Q_6, Q_8) \\ dQ_5/dt = f_5'\,(Q_4, Q_5, Q_6, Q_8) \\ dQ_6/dt = f_6'\,(Q_4, Q_5, Q_6, Q_8) \\ dQ_8/dt = f_8\,(Q_4, Q_5, Q_6, Q_8) \end{cases} \tag{7}$$

$$S^3 : \begin{cases} dQ_6/dt = f_6''\,(Q_6, Q_7, Q_9) \\ dQ_7/dt = f_7'\,(Q_6, Q_7, Q_9) \\ dQ_9/dt = f_9\,(Q_6, Q_7, Q_9) \end{cases} \tag{8}$$

The associations between the systems and their state variables are then:

$$\begin{cases} S^1 : (Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ S^2 : (Q_4, Q_5, Q_6, Q_8) \\ S^3 : (Q_6, Q_7, Q_9) \end{cases} \tag{9}$$

Because of their simultaneous occurrence in different systems, the common state variables must simultaneously behave as components of different systems due to the constraints as in Eqs. (10), (11), and (12).

$$\begin{cases} dQ_5/dt = f_5\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_5/dt = f_5'\,(Q_4, Q_5, Q_6, Q_8) \end{cases} \tag{10}$$

$$\begin{cases} dQ_6/dt = f_6\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_6/dt = f_6'\,(Q_4, Q_5, Q_6, Q_8) \\ dQ_6/dt = f_6''\,(Q_6, Q_7, Q_9) \end{cases} \tag{11}$$

$$\begin{cases} dQ_7/dt = f_7\,(Q_1, Q_2, Q_3, Q_5, Q_6, Q_7) \\ dQ_7/dt = f_7'\,(Q_6, Q_7, Q_9) \end{cases} \tag{12}$$

The validity regime is quasi-ruled by the irregular and partial simultaneous occurring of validity of equations of the type (6), (7), (8). We notice that the equations ruling simultaneous variables introduce constraints that lower the number of degrees of freedom of the original description. **The reduction of the number of degrees of freedom has significant implications on the stability of the combined motions of collective systems. The increasing of the dimensionality of the phase space also increases the number of ways in which an equilibrium state can become unstable. On the contrary, the reduction in the number of degrees of freedom increases stability. In general, a collective system is, in principle, more stable than its local constituent parts, and this stability is, in turn, granted only by the defining constraints**. **The reduction of the number of degrees of freedom**

**should be intended to maintain coherence by the collective behavior when some more suitable constraints are active, such as ergodicity.**

In this regard, we may consider why some collective systems, such as two-dimensional flocks, seem to violate well-known theorems of physics, such as the Mermin-Wagner theorem, stating that a stable two-dimensional configuration cannot exist (Mermin & Wagner, 1966). This is also related to consider *order parameters in Synergetics*, an approach that reduces the number of degrees of freedom to only a few parameters (Haken, 1987, p. 425). Furthermore, more sophisticated versions of the previous mechanism relate to considering crossed multiplicities of validity regimes not expressible as above with systems of structurally, functionally crossed ordinary differential equations. For example, when different validity regimes give quasi-ness interesting, in variable ways, different domains of entities, cross or superimpose in any way. This is also the case, for instance, when behavioral paths of elements display chaotic, ergodic, correlated, networked features identifying related corresponding validity regimes. However, it is possible to have cases of multiple regimes because, for instance, of the multiple roles of components, e.g., ecosystems, inferable as multiple interaction mechanisms, and displaying multiple properties.

The case considered in this section should be intended as an elementary example of validity regimes establishing complex systems where various superimposed processes of emergence occur, allowing dynamics of acquisition, losing, and resumption of coherences in various percentages elaborated in the following Sects. 4 and 5.

*Dealing with complexity, we face the limited power of analytical formalizations finalized to completeness in representations (*Minati & Pessa, 2018*, pp. 192–198). It is necessary to represent incompleteness, quasi-ness, and levels of coherence (coherence of local, sub-coherences, see* Fig. 1*) as properties of validity regimes and not intended as a detrimental pernicious inaccuracy.*

*It is matter to use approaches, methods, and tools based, for instance, on (deep) learning in neural networks, profiling techniques, networks, and the mathematics of the so-called naturally-inspired computations such as ant algorithms, artificial immune system algorithms, bacterial foraging algorithms, chemically inspired algorithms, developmental and grammatical computing, evolutionary computing, genetic algorithms, genetic programming, genetic regulatory networks, grammar and genetic programming, grammar-based genetic programming, grammatical evolution, neuro-evolution algorithms, particle swarm algorithms, plant-inspired algorithms, and quantum-inspired evolutionary algorithms* (Brabazon et al., 2015; Mac Lennan, 2004; Liu et al., 2020).

**Fig. 1** A coherent flock of multiple, variable coherent multiple intersecting and interfering sub-flocks

## 4 The Nature of Edges, Interfaces, and Transience

In this section, we consider the dynamics of validity regimes, particularly their edging, interfacing, and transience processes representing the dynamics of the complexity of collective systems.

### *4.1 Edges*

In conceptual correspondence with propagation phenomena in physics and the so-called moving boundary problems (Chakrabarti & Brebbia, 2007), we consider edges of validity regimes having multiple natures. Examples of related problems considered in physics are the so-called *Stefan problem* (Meirmanov, 1992), describing the melting of a semi-infinite sheet of ice when the surface maintained at a temperature greater than the melting temperature (Minati & Pessa, 2018, pp. 50 and 71), processes of phase separation occurring within a binary fluid (the two components separate and give rise to two spatial domains), and general complex models of phase change.

A general understanding of the concept of *edge* and its topological content, relates to the role of separating, demarcating, distinguishing absolutely or more probably, in increasing—decreasing, oscillatory or fuzzy ways. Generic examples of edges are frontiers of generic and volcanic eruptions, studying how they spread, in case mix in the environment, such as seismic waves. The related fuzziness is considered, for instance, in approaches for image processing and classification (Liang & Looney, 2003).

*We consider here the edges of validity regimes.*

We consider the cases where edging is between validity regimes of processes of emergence involving the same components, such as populations of interacting clocks (Mikhailov & Calenbuhr, 2002, p. 127) and logistic maps (Mikhailov & Calenbuhr, 2002, p. 155; Minati & Licata, 2012); multiple local coherences in collective behaviors such as in multiple flocks (see Fig. 1) and between levels of emergence (Ballerini et al., 2007; Minati & Pessa, 2018, pp. 253 and 286).

**Edges are understandable as topological places where changes between almost two validity regimes occur, where one can be the empty validity regime. Edges suppose *some*, probably irregular, multiple occurrences of properties, interactions, and interferences between non-equivalent validity regimes.**

Edges are supposed provided with shading, a fuzzy temporary predominance of single or more validity regimes, environmental compositions, and propagations from the single or multiple validity regimes of the corresponding complex system. We consider the edges of interaction mechanisms and properties.

Predominance is supposed to be dynamic when considering the evolution of the edges' spatial dimensions, the number of affected components, and topological properties. It is then possible to consider the *dynamics of edges*. Edges may also be intended to occur inside (different topologies are possible) complex systems, separating validity regimes, such as the different emergence processes when edges act as (fuzzy) borders of the different processes of emergence. The interest is in the *dynamics* of edges establishing *waves complexity* corresponding to the expansion, retreat, and combinations of the different validity regimes for processes of emergence.

*Edges are intended as topological places where different validity regimes exist, for instance, situations of simultaneity, superposition, composition up to the increasing formations of predominance, reduction of validity, or their variable combination.*

We think of extreme cases where the dimensionalities and temporalities are such that there are only edges when no significant temporal and dimensional domains are established, having some stability and predominance. In such cases, we may say that there are only *waves of complexity*. *Waves of complexity are supposed to occur in multiple, subsequent processes of emergence, specifying radical emergence* understandable, from this point of view as a *turbulent emergence* of waves of complexity (see Sect. 4.3). Turbulent regimes may be intended, constituted of instable, having low-recurrent and an irrelevant subsequent irregular multitude of edges.

When dealing with the ‘constant’ homogeneity of validity regimes and zones valid for the entire system, we may say that no edges and transformation processes are diluted over the whole validity regime. **We underlie how it is, however, matter of simplification, *reversely engineered* or *abductively inferred* models of interacting elements’ phenomenological behavioral properties of which different representations can be given, in a non-univocal way. With the multiple roles of interacting elements, the latter establishes the logical openness and multiplicity of the complexity of collective behaviors and edges**.

## *4.2 Interfaces*

Interfacing (Hookway, 2014; Longo, 2019) is intended as an activity and edging role. It is supposed active when the intermediated flux of information through edges is flowing, in some way, processed and *supervised*. Examples are active selection processes (not just filtering), progressive replacement, regulations, acquisition of oscillatory balances, and weighting.

Edges are the places where the interfacing occurs, whose elements operate by interfacing. However, we should consider how spatial and temporal adjacency of boundaries, not absolute edges (closedness is partial), often makes interfacing inescapable or even unwilled role. On the other side, we may ask what is not an interface? No interfacing may be considered coincident with total isolation (when edges are reduced to impermeable barriers) or total equivalence, homogeneity. Furthermore, interfacing may be unidirectional, bidirectional, and their combinations.

An interface should be not intended reduced to *separation,* such as for the properties of materials, e.g., immiscible water—oil, where clear boundaries are distinguished. An interface is considered empty when there are no intermediate values between belonging—non-belonging to validity regimes, e.g., second-order phase transitions.

Interfaces of this kind may be supposed to exist where there is no complexity, no emergence. More specifically, where there is a single all—comprehensive homogeneity, a single structural regime, cardinality, and where interfaces are vanishing as in cases of continuous dilution or in the case of processes of aggregation. On the other hand, we may have no or temporary interfaces in case of dilution, fading (vanishing until to pass the Avogadro number), or aggregation processes, concentration leading to the subsequent creation of separations or interfaces. However, the interfaces should be understood as an *active between* influencing interaction or even exchanging energy and information. We should notice that interfacing is not given by introducing *deformations* or *noise* in processes of interaction.

Cases of interfacing include transformation and adaptation, conversion, and *negotiations* between macro and micro in processes of emergence, as in the ‘middle way’ (Laughlin et al., 2000). Interfaces and interfacing, widely used in several disciplines and applications; in some cases, interfacing may be considered reduced, degenerated to active bordering for protection and defense, intended as sheltering, regulation, and generate antagonistic vanquishing reactions. Furthermore, interfaces may be supposed to have active roles. They can be considered, for instance, as transforming a representation into other ones eventually non-one-to-one correspondent, and non-equivalent. The process of translating is a related example of interfacing. Interfacing occurs in computer science, telecommunications, and engineering (Artemiadis, 2014). Other cases relate to interfacing as ordering, selecting information, and transforming signals, for instance, from acoustic to electrical.

Interfaces should be intended as variable and multiple, eventually combined and occurring in various ways, such as adaptive or context-sensitive, having learning features, and predictive. This is the case for human-machine interfaces that adapt, for example, through machine learning, themselves to their user's characteristics.

A more general systemic understanding of the interface concept may relate to considering them as the *place* where processes of balancing, negotiation, reformulation, representation, selection, transformation, transition, and translation occur *among non-equivalences*. Interfacing is a typical property of most complex systems relating to single collectively interacting entities and their temporal sub-communities having specific temporal dynamics, establishing waves of complexity (see Sect. 4.3), such as in collective behaviors and within ecosystems.

**Interfacing is performed by active edging when having not the only purpose to actively close, separate (usually performed passively as a closed border). However, the interface may *decide* to close.**

**Interfaces may not necessarily be active in some cases. Still, they can be performed in other ways, by insertion of *diffusive* processing through artificial devices, partial barriers, and parallel information processing imposed artificially, such as external classification, ordering, and selection. In such cases, we may speak, however, of *virtual edging*.**

In sum, edges and interfaces have a dual nature: active edges perform roles of interfaces, and interfaces perform roles of edges, having, however, both variable, often context-sensitive predominance.

In the following, with such double nature in mind, the terms 'edge' and 'interface' are considered essentially equivalent and used separately when one role is considered more predominant.

## *4.3 Transience*

Conceptually, interfacing, an active role of edges, occurs between validity regimes. Transience may be intended as interfacing, performing the on-going establishment of one regime's predominance over the other until its substantial replacement. It may occur in different ways, be inversed, oscillatory, or unstable.

Transience is supposed to also occur *inside* a validity regime when internal transience is activated by external perturbations inducing, for instance, meta-stability or by internal processes of acquisition of properties, e.g., multiple emergence and self-organization. Transience from a validity regime to another one may occur in several ways, context-dependent. Transience between regimes of validity may be intended, as a mechanism of complex systems dynamics, dynamics having multiple, as far as possible to distinguish aspects, e.g., altitudes, density, distances, energy, and speed.

**Transience may be intended as a process consisting of establishing (local, multiple, temporary, or unique) convergences and possibly re-establishing the previous predominance in interfacing.**

We will consider in the following how transience is intended performed by interfaces**.** While interfacing has no privileged dynamics and directions, *transience* is intended mostly as unidirectional, however, convergent interfacing. In non-elementary cases, the change of a validity regime occurs through *multiple interfaces* corresponding to *particular dynamics of edges*. e.g., constitutive, expansive, recessive, combinative dynamics for specifying edging.

Examples of transience with no interfacing, i.e., the simultaneous variation of the validity regime, include metastability and second-order phase transitions such as paramagnetic to magnetic, and the acquisition of properties, such as superfluidity and superconductivity. However, the process of transience is interesting as it consists of the change in the predominance of properties.

It is a matter of predominance of properties, characterizing the 'transformation' from a validity regime into another. We may consider the modalities with information and materiality flow through the interfaces, e.g., intermediated by active selections, processing, parasitic absorption of energy, reducing degrees of freedom, transformations, and weighing. In this regard, we mention the concept of structural dynamics (see Sect. 4.3.1) in complex systems intended as change and multiplicity of roles of elements, structures between them (Minati & Pessa, 2018, pp. 87–90, 102–130), and as dynamics of edges and validity regimes.

Examples of processes of the transience of interest here are between generic validity regimes such as:

- Order-disorder with variable predominance and homogeneity;
- Different processes of emergence (the system is acquiring multiple or losing emergent properties);
- Openness and closedness (when a system is closing or opening, with variable predominance and homogeneity);
- Remote synchronization acquisition. (the system acquires or loses synchronization) (Minati, 2015).

and

- Establishing multiple emergences inside the validity regime of a specific process of emergence, as in collective behaviors acquiring multiple different properties, e.g., topological in highly dynamical flocks or waves of complexity. For instance, as manifested by giant honeybees' social waves (Kastberger et al., 2010) and in flocks of starlings, see Fig. 1 and Ballerini et al. (2007).

### 4.3.1 Transience as Converging Structural Dynamics

It is possible to consider continuous partial converging interfacing cases, with a significant increase in a validity regime's predominance. In such cases, transience occurs. Otherwise, it is possible to consider instances of continuous partial non-convergent interfacing, with non-significative increasing of a validity regime's predominance. In such cases, processes of transience are continuously in progress and have no definitive transient consolidation.

When considering transience in system structural dynamics processes, such a dynamic converges, as the acquisition of synchronicity and coherence, starting from the incomplete, tentative *initial conditions* of a self-establishing, quasi-convergent process. This describes *spontaneous synchronizations* such as in applauding, for objects on vibrating surfaces, and the emissions initially unsynchronized of flashes of light (bioluminescence) by communities of fireflies until a specific synchronization becomes predominant and iterated. However, such convergences are rarely a transformation process with the character of regularity, continuity, but rather an irregular sequence as for quasi-ness. Structural dynamics are represented ideally by structural interactions, such as $f_n$ and $g_n$ in Eqs. (10), (11), and (12), specified by processes of acquisition, change, combination, and loss of structures such as occurring for phase transitions and networks.

Examples of complex structural dynamics are given by the cytoskeleton dynamics, consisting, within the cell cytoplasm, of a network of protein fibers and characterized by structural dynamics, as its parts are continuously destroyed, renewed or newly created; self-organized and emergent collective behaviors (see Footnote 5), and social networks.

**This opens the possibility to consider the structural dynamics of complex systems as dynamics of the transience of interfacing edges having correspondence with the dynamics of multiple, embedded processes of emergence (see Sect. 5).**

Interesting variables to be considered are the delay in spreading through of the edges, their interfacing time, and the related structural dynamics. Furthermore, in mathematics the whatever transience of the structural dynamics may be related to the acquisition, intended as the transition from computation to acquired *computational properties of computations in progress* (emergent computation), such as for Artificial Neural Networks and Cellular Automata, rather than to formal properties of the representative equations such as for classic mathematics (Licata & Minati, 2016). **In such cases the computed is not deducible from the computing mechanism.**

Correspondingly, in mathematics, the end of the so-called Bourbaki program (1935–1998) relying on abstract definitions and axioms and finalized to a *completely* self-contained treatment of the core areas of mathematics was a manifestation of the decreasing effectiveness and role of classical formalist mathematics relying on abstract definitions and axioms.

**It is a matter to consider the *autonomy* (in acquiring properties) of computational processing in progress and devices suitable to represent and constitute *per se* the structural autonomy of complex systems continuously acquiring**

**Table 2** Edges, interfaces, transience, structural dynamics, waves of complexity

| |
|---|
| Edges |
| understandable as topological places where changes between almost two validity regimes occur, where one can be the empty validity regime. Same or few interaction mechanisms and ergodic or network properties predominate. |
| Interfaces |
| interfacing is intended as activity, role of edges as intermediating between validity regimes. |
| *Active edges and interfaces are mostly equivalent terms distinguished when one role prevails.* |
| Transience |
| from a validity regime to another one, is supposed to occur as predominant directional interfacing and *inside* a validity regime, e.g., the establishing of multiple emergences inside the validity regime of a specific process of emergence. |
| Structural dynamics |
| changes of structure among elements, e. g. sequences of phase-transitions-like changes as in coherent sequences of local, sub-emergences and the cytoplasm. |
| Waves of complexity |
| intended as sequences of interfacing edges, transience, and structural changes having significant local coherences, e.g., sequences of processes of emergence involving same components such as highly dynamic multiple flocks acquiring topologically different forms. |

**properties**. For example, we mention the sub-symbolic processing of artificial neural networks (ANNs) and their deep machine learning through recurrent neural networks (RNNs) (Bianchi et al., 2017).

The autonomy of computational approaches and devices should be intended as suitable for modeling complexity, for example, in physics and chemistry, and respectful and appropriate for social, medical, environmental, and even economic phenomenological processes. The consideration of only complete formal approaches and general characterizing formal properties. e.g., chaoticity and ergodicity, are effective when considering only specific aspects and assuming their sufficiency to explain and model the global structural dynamics.

We conclude this section by summarizing (see Table 2) how in *highly* complex systems, i.e., systems in which *multiple processes emergence occur*, the corresponding dynamic interfacing edges of validity regimes and transience should be considered as waves of complexity, representing their complexity dynamics.

The fields of Statistical Physics and Thermodynamics are concerned with physical systems containing a large number of particles. Examples include gases, liquids, solids, and photon gases. In fact, most systems are large; isolated particles rarely occur.

## 5 Dynamics of Complex Systems as Transience Among Validity Regimes

In this section, we further elaborate on the dynamics of change *between* validity regimes. In this regard, we mention how several researches deal with transition processes *between* phases such as non-system and system, disorder and order, non-chaos and chaos, living and non-living, and in general between different phases of matter studied in the physics of phase transitions. For instance, this relates to emergence processes, which is understood as the occurrence of possibly multiple simultaneous sequences of self-organization processes or local emergence (for their distinction, see Footnote 5). When the corresponding acquired dynamic structures are globally coherent (see, for instance, the theory of 'dual evolution' for adaptive systems, as in Paperin et al., 2011).

We may relate to the processes of emergence, like in physics, as associated with sequences of phase transitions (Sachdev, 2011; Solé, 2011) and the *dynamics of self-organization and emergence* (Minati & Pessa, 2018, pp. 80–87). Such sequences require the occurrence of at least two phases, such as one following the other *after* the transition, physically not equivalent. It means that it is impossible to find a transformation that reduces one phase's physical description to one of the other.

Properties of the transience may specify the *nature of a systems' complexity* when having multiple regimes, like given by chaos, correlation, ergodicity, networking, and, of interest here, by their eventual combinations. A related example case takes place when considering *multiple dynamical attractors* and properties of their dynamics (Kauffman, 2011; Scarpetta et al., 2008) in superimposed abstract spaces of multiple attractors. Each space, considered as corresponding to different regimes.

We may also consider cases of phases, fractals, and symmetries:

- *Multiple phases* describe elements of the system involved in different phases. In such a case, the dynamics related to the changes of phases involving in multiple ways the elements and then establishing eventual multiple roles of same elements (Brovchenko et al., 2005; Brovchenko & Oleinikova, 2008).
- *Multiple fractals* when multiple fractal rules occur and the same segments or surfaces belong to multiple, different fractals. Dynamics relates to the changes in rules and eventual multiple roles of segments and surfaces (Chen, 2014; Harte, 2001).
- *General multiple symmetries* describe when multiple symmetries are established in connected networks of nodes. Same nodes may belong to different networks (Nicosia et al., 2013) having different symmetries. Dynamics relates to the change in structures and multiple roles of the elements-nodes establishing multiple symmetries (McClain, 2008; Pessa, 1988). "A way of looking at this situation is to reinterpret the observed deviation from the exact symmetry as a phenomenological distortion or rearrangement of the basic symmetry ... The crucial problem one has to face in the recognition of a symmetry is, then, the intrinsic two-level description of Nature . . . This two-level description of Nature

> was soon recognized in Quantum Field Theory (QFT) as the duality between fields and particles." (Blasone et al., 2011, p. 1).

We conclude by stressing how transience of complex systems is intended as consistent, on-going betweenness, interfacing, and waves of complexity, mentioning two related issues:

1. *Topology of complexity*

   The study of validity regimes, interfacing edges, and transience may be approached in the conceptual framework of the topology of complexity, considering compatibilities and incompatibilities of configurations and their eligible compositions and transformations. Such topology of complexity, should not be considered as an exhausted topic by properties of complex networks in network science. The validity regimes' topologies and interfacing edges should be considered as constituting the admissible topologies of transience.
2. *Meta-structural transience*

   The roles and dynamics of equivalences should be considered for the validity regimes, representable by setting meta-structural representation levels. Here we may mention two cases:

   (a) *Meta-structural transience* intended to occur when the transience relates to the acquisition, change, or loss of a specific meta-structural property,[8] in short, properties of clusters;

   (b) *Transient between meta-structural regimes of validity when meta-structural properties are still maintained, but in different ways*, through different clustering and parameterizations.

## 6 Further Research

First of all, further research should focus on suitable new approaches to formalize (Minati & Pessa, 2018, pp. 187–219), models and simulations of the approaches proposed above, particularly in Sect. 3.1 and related to the concepts of validity regime, edge, interface, transience, and waves of complexity in complex systems.

We mention in the following some topics of further research such as:

- Topological dynamics of validity regimes and edges;
- Superimpositions of validity regimes and edges;
- Intra-dynamics of validity regimes, their losing of validity and resumption of validity;

---

[8] Structures among cluster properties are considered *meta* because of the incompleteness and variability of clusters, e.g., in number of components over time; in their non-regular, non-iterative occurrence; and in number of shared elements belonging to them over time (Minati & Licata, 2012, 2013, 2015; Minati et al., 2013; Minati & Pessa, 2018, pp. 102–129).

- Validity regimes in ecosystems and social systems;
- Validity regimes vs. fields.

## 7 Conclusions

This section's title sounds bitter since the chapter concentrated most on the uncompleted common scientific and cultural journey I have had in progress with Eliano. I naively presented only a small percentage of ideas, conjectures, and proposals that I would have liked to submit to him.

I never wanted to be here to write this chapter. I am experiencing the shame of survivors. We must live the time available and ensure we deserve it, to play our current role with the utmost dedication; this is our duty. How did we deserve to have Eliano as a friend, companion, and colleague?

We delineated in several publications the continuous trade-off between coherence—incoherence, incompleteness—quasi-ness, levels of emergence, the infiniteness of betweenness, and collapsing mechanisms. The essence is that it is a matter to acquire open approaches suitable to induce and orient the complexity, *theoretically* irreducible, non-zippable into analytical *complete* or *completable* representations.

This chapter's conclusions may bring to the attention of the systemic and scientific community the need to acquire a new and extended understanding of the complexity. Suitable to overcome the approach based on the searching for appropriated representations and rules assumed theoretically able to represent and illustrate complex phenomena *exhaustively*. The usual assumption is that incompleteness is incompatible, at the most momentarily tolerable, with scientific approaches (exceptions are, for example, so-called grey systems, fuzzy logic, probabilistic and statistic approaches). Eliano and I introduced in the literature contributions, even mentioned in this chapter, where incompleteness, we considered theoretically necessary for radical emergence processes. This theoretically changes the approaches to be considered for modeling, representing, and dealing with complexity. The related key-concepts introduced here are the ones of validity regimes and their quasi-ness. We mentioned an elementary example, however, possibly significant since dealing with analytical representations with ordinary differential equations—completeness context—rather than with sub-symbolic or network cases.

From physics, we have extrapolated to the validity regimes the concepts of edges, fields, interfaces, transience, allowing us to consider transience as converging structural dynamics and dynamics of complex systems as transience among validity regimes. In this conceptual framework, we introduced the idea of *waves of complexity* corresponding to the expansion, retreat, and combinations of the different validity regimes for processes of emergence. This may allow future, more appropriate models and representations of the dynamics of complexity, models in which aspects of autonomy and incompleteness are theoretically essential and ethically respected without assuming complete prescribability and full theoretical decidability for granted.

Thank you, Eliano, for all that you taught me, with the simplicity of who has climbed the highest peaks of humanity, science, and of course mountaineering. We have all certainly lost a lot.

## References

Artemiadis, P. (2014). *Neuro-robotics: From brain machine interfaces to rehabilitation robotics*. Springer.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., & Zdravkovic, V. (2007). Interaction ruling animal collective behaviour depends on topological rather than metric distance: Evidence from a field study. *Proceeding of the National Academy of Sciences of the United States of America, 105*(4), 1232–1237.

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). *Recurrent neural networks for short-term load forecasting: An overview and comparative analysis*. Springer.

Blasone, M., Jizba, P., & Vitiello, G. (2011). *Quantum field theory and its macroscopic manifestations*. Imperial College Press.

Brabazon, A., O'Neill, M., & McGarraghy, S. (2015). *Natural computing algorithms*. Springer.

Brovchenko, I., Geiger, A., & Oleinikova, A. (2005). Liquid-liquid phase transitions in supercooled water studied by computer simulations of various water models. *The Journal of Chemical Physics, 123*(4), 44515.

Brovchenko, I., & Oleinikova, A. (2008). Multiple phases of liquid water. *ChemPhysChem, 9*(18), 2660–2675.

Chakrabarti, S. K., & Brebbia, C. A. (2007). *Fluid structure interaction and moving boundary problems IV*. WIT Press.

Chen, Y. (2014). Multifractals of central place systems: Models, dimension spectrums, and empirical analysis. *Physica A, 402*, 266–282.

Gühne, O., & Toth, G. (2009). Entanglement detection. *Physics Reports, 474*, 1–75.

Haken, H. (1987). Synergetics: An approach to self-organization. In F. E. Yates (Ed.), *Self-organizing systems: The emergence of order* (pp. 417–434). Plenum.

Harte, D. (2001). *Multifractals: Theory and applications*. Chapman and Hall/CRC Press.

Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J. T., & Ward, A. J. W. (2011). Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences of the United States of America, 108*(46), 18726–18731. https://doi.org/10.1073/pnas.1109355108

Hookway, B. (2014). *Interface*. MIT Press.

Javanmardi, E., Liu, S., & Xie, N. (2020). Exploring the philosophical paradigm of grey systems theory as a postmodern theory. *Foundations of Science, 25*, 905–925. https://doi.org/10.1007/s10699-019-09640-5

Kastberger, G., Weihmann, F., & Hoetzl, T. (2010). Complex social waves of giant honeybees provoked by a dummy wasp support the special-agent hypothesis. *Communicative & Integrative Biology, 3*(2), 179–180.

Kauffman, S. A. (2011). Approaches to the origin of life on earth. *Life, 1*(1), 34–48.

Laughlin, R. B., Pines, D., Schmalian, J., Stojkovic, B. P., & Wolynes, P. (2000). The middle way. *Proceedings of the National Academy of Sciences of the United States of America, 97*(1), 32–37.

Liang, L. R., & Looney, C. G. (2003). Competitive fuzzy edge detection. *Applied Soft Computing, 3*(2), 123–137.

Licata, I., & Minati, G. (2016). Emergence, computation and the freedom degree loss information principle in complex systems. *Foundations of Science, 21*(3), 1–19.

Liu, Y., Wang, L., Zhao, L., & Yu, Z. (Eds.). (2020). *Advances in natural computation, fuzzy systems and knowledge discovery*. Springer Nature.

Liu, S., & Yang, Y. (2012). A brief introduction to grey systems theory. *Grey Systems: Theory and Application, 2*(2), 89–104. https://doi.org/10.1108/20439371211260081

Longo, G. (2011). Reflections on concrete incompleteness. *Philosophia Mathematica, 19*, 255–280.

Longo, G. (2019). Interfaces of incompleteness. In G. Minati, M. R. Abram, & E. Pessa (Eds.), *Systemics of incompleteness and quasi-systems* (pp. 3–55). Springer.

Mac Lennan, B. J. (2004). Natural computation and non-Turing models of computation. *Theoretical Computer Science, 317*(1–3), 115–145.

McClain, W. M. (2008). *Multiple symmetries in symmetry theory in molecular physics with mathematica* (pp. 549–563). Springer.

Meirmanov, A. M. (1992). *The Stefan problem*. De Gruyter.

Mermin, N. D., & Wagner, H. (1966). Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Physical Review Letters, 17*, 1133–1136.

Mikhailov, A. S., & Calenbuhr, V. (2002). *From cells to societies. Models of complex coherent actions*. Springer.

Minati, L. (2015). Remote synchronization of amplitudes across an experimental ring of non-linear oscillators. *Chaos, 25*, 123107–123112.

Minati, G. (2016a). Knowledge to manage the knowledge society: The concept of *theoretical incompleteness*. *Systems, 4*(3), 1–19. http://www.mdpi.com/2079-8954/4/3/26

Minati, G. (2016b). General System(s) Theory 2.0: A brief outline. In G. Minati, M. Abram, & E. Pessa (Eds.), *Towards a post-Bertalanffy systemics* (pp. 211–219). Springer.

Minati, G. (2018). The *non-systemic usages* of systems as reductionism. Quasi-systems and quasi-systemics. *Systems, 6*(3), 28. http://www.mdpi.com/2079-8954/6/3/28

Minati, G. (2019a). Non-classical systemics of quasi-coherence: From formal properties to representations of generative mechanisms. A conceptual introduction to a paradigm-shift. *Systems, 7*(4), 51. https://www.mdpi.com/2079-8954/7/4/51

Minati, G. (2019b). Phenomenological structural dynamics of emergence: An overview of how emergence emerges. In U. U. Lucia (Ed.), *The systemic turn in human and natural sciences. A rock in the pond* (pp. 1–39). Springer.

Minati, G. (2019c). On some open issues in systemics. In G. Minati, M. Abram, & E. Pessa (Eds.), *Systemics of incompleteness and quasi-systems* (pp. 317–323). Springer.

Minati, G., Abram, M., & Pessa, G. (Eds.). (2019). *Systemics of incompleteness and quasi-systems*. Springer.

Minati, G., & Licata, I. (2012). Meta-structural properties in *collective behaviours*. *The International Journal of General Systems, 41*(3), 289–311.

Minati, G., & Licata, I. (2013). Emergence as mesoscopic coherence. *Systems, 1*(4), 50–65. http://www.mdpi.com/2079-8954/1/4/50

Minati, G., & Licata, I. (2015). Meta-structures as multidynamics systems approach. Some introductory outlines. *Journal on Systemics, Cybernetics and Informatics (JSCI), 13*(4), 35–38. http://www.iiisci.org/journal/sci/issue.asp?is=ISS1504

Minati, G., Licata, I., & Pessa, E. (2013). Meta-structures: The search of coherence in collective behaviours (without physics). In A. Graudenzi, G. Caravagna, G. Mauri, & M. Antoniotti (Eds.), Proceedings Wivace 2013—Italian Workshop on Artificial Life and Evolutionary Computation (Wivace 2013), Milan, Italy, July 1-2, 2013, Electronic Proceedings in Theoretical Computer Science 130 (pp. 35–42). http://rvg.web.cse.unsw.edu.au/eptcs/paper.cgi?Wivace2013.6.

Minati, G., Penna, M. P., & Pessa, E. (1996). Towards a general theory of logically open systems. In E. Pessa, M. P. Penna, & A. Montesanto (Eds.), *Proceedings of the 3rd Systems Science European Congress* (pp. 957–960). Kappa.

Minati, G., Penna, M. P., & Pessa, E. (1998). Thermodynamic and logical openness in general systems. *Systems Research and Behavioral Science, 15*(3), 131–145.

Minati, G., & Pessa, E. (2006). *Collective beings*. Springer.

Minati, G., & Pessa, E. (2018). *From collective beings to quasi-systems*. Springer.

Nicosia, V., Bianconi, G., Latora, V., & Barthelemy, M. (2013). Growing multiplex networks. *Physical Review Letters, 111*(058701), 1–5.

Paperin, G., Green, D. G., & Sadedin, S. (2011). Dual-phase evolution in complex adaptive systems. *Interface, 8*, 609–629.

Pessa, E. (1988). Symmetry breaking in neural nets. *Biological Cybernetics, 59*, 277–281.

Preparata, G. (2002). *Introduction to a realistic quantum physics*. World Scientific.

Reynolds, C. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*, 25–34.

Sachdev, S. (2011). *Quantum phase transitions*. Cambridge University Press.

Scarpetta, S., Yoshioka, M., & Marinaro, M. (2008). Encoding and replay of dynamic attractors with multiple frequencies: Analysis of a STDP based learning rule. In M. Marinaro, S. Scarpetta, & Y. Yamaguchi (Eds.), *Dynamic brain—From neural spikes to behaviors* (Vol. 5286, pp. 38–60). Springer.

Slowik, M. (2012). *Metastability in stochastic dynamics*. Südwestdeutscher Verlag Fur Hochschulschriften AG.

Solé, R. V. (2011). *Phase transitions*. Princeton University Press.

Vicsek, T., & Zafeiris, A. (2012). Collective motion. *Physics Reports, 517*, 71–140.

von Bertalanffy, L. (1968). *General system theory. Development, applications*. George Braziller.

Zadeh, L. A., Klir, G. J., & Yuan, B. (Eds.). (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: Selected papers by Lotfi A. Zadeh*. World Scientific.

# The Use of tDCS Combined with CET Training for the Treatment of Pathological Dependence

**Natale Salvatore Bonfiglio, Roberta Renati, and Maria Pietronilla Penna**

**Abstract** Studies on pathological addictions have shown how the need and the search for the substance are stimulated by environmental situations linked to the substance (trigger). This condition is determinant for the state of craving (Bonfiglio et al., Addict Behav Rep 9:100172, 2019). Craving is considered as a conditioning response linked to the search for the substance and determined by the subject's impulsiveness and inability to control himself. Several studies have shown how it is possible to reduce the need for craving and impulsivity through neurostimulation with tDCS (Transcranial direct-current stimulation). Other studies have obtained promising results in this area through the cue exposure paradigm (CET), which consists of presenting the subject with a series of trigger stimuli, which recall the substance, desensitizing its effect and increasing self-control. This work presents an example of a treatment that uses neurostimulation with tDCS together with the cue exposure paradigm on 10 subjects with sham tDCS and 10 with active tDCS, compared with 20 control subjects. After 10 sessions of neurostimulation with active tDCS and sham and cue exposure, the results seem to confirm the hypothesis of a reduction in craving levels and ability to resist for the condition with active tDCS and partially for the condition with sham tDCS. There were no improvements in impulsivity levels. The proposed treatment, despite the partial results, shows many potential, above all due to the possibility of a certain autonomy of use—in the absence of an operator—which goes against the current progress in the field of telemedicine and treatment through a remotely planned and supervised program.

---

N. S. Bonfiglio (✉) · R. Renati
University of Pavia, Pavia, Italy

Noah srl, Pavia, Italy
e-mail: salvo.bonfiglio@unipv.it; s.bonfiglio@noahealth.it; roberta.renati@unipv.it; r.renati@noahealth.it

M. P. Penna
Department of Pedagogy, Psychology, Philosophy, University of Cagliari, Cagliari, Italy
e-mail: penna@unica.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_9

**Keywords** Addiction · Craving · Cue exposure · Dependence · Impulsivity · Neurostimultation · Self-control · tDCS · Training

## 1 Introduction

Pathological dependence is a challenge for the World Health Organisation as it tends to become a chronic problem for addicted people and results in high costs for society. Pathological addiction is difficult to treat, especially because it very often involves chronic relapse, despite acute detox and withdrawal.

Many treatments are based on managing the patient's craving (which is chiefly to blame for relapse) and impulsivity by focusing on the direct effect that gratification has in this sense (Johnson, 2008; Addolorato et al., 2006; De Mulder & Dom, 2012).

Impulsivity is directly linked to a craving response and it is one of the main factors responsible for relapse (Wrase et al., 2008; Koob & Volkow, 2010); additionally, being aware of their own craving and the o*t*her causes that lead to relapse do not deter addicts from maintaining abstinence (Tiffany et al., 2000). This is likely to happen because the process leading to relapse is essentially automatic and uncontrollable in each individual, as explained by George and Koob (2011) via his three-stage cycle of addiction. As a matter of fact, pathological dependence refers to the final stage of a process that starts with an (often recreational) usage of a substance leading to dependant behaviour that is driven by impulsivity and obsessive compulsivity towards that substance (e.g. alcohol, cocaine, etc.) or the object of addiction (e.g. gambling, work, etc.). Usage becomes pathological when a person "loses control" over their drug-seeking and intake (Koob & Le Moal, 2008). From this standpoint, addiction is characterised by (a) compulsive drug-seeking and drug-taking; (b) loss of control over drug-taking leading to (c) negative emotional state (e.g. dysphoria, anxiety, irritability, etc.) because obtaining drugs becomes difficult or it is impaired (Koob & Le Moal, 2008).

An important aspect of experimental research is that it attempts to understand how individuals move from controlled drug use to the compulsive uncontrolled state that characterises addiction (Koob & Le Moal, 2008), which also includes neural and neurobiological mechanisms connected to dependence itself. The hypothesis that has been exhaustively verified suggests addiction results from a process that activates natural motivational systems and their related neural circuitry, such as the reward/gratification system (Koob & Le Moal, 2008); also, the dopamine system seems to be responsible for addiction (Kienast et al., 2013). Individuals suffering from pathological dependence display a dysregulation of the dopamine system, which leads to an increasing loss of motivation for natural rewards (such as food) and an increased interest in the drug as a the main and most important source of strength (Heinz et al., 2009).

Many brain regions are involved in the gratification/reward system and dopamine circuit, such as the dorsolateral-pre-frontal-cortex (DLPCF), the nucleus accumbens and the ventral tegmental area (VTA) (Bechara, 2005). Besides, compulsivity seems to be a type of behaviour deriving from the dysregulation of the dopamine system caused by the activation of the hypothalamic-pituitary-adrenal (HPA) axis as a stress response.

Recent neuroimaging studies have shown that the left dorsolateral-pre-frontal-cortex (DLPC) is where craving, as a pathology, is activated; this brain region also plays an essential role in craving regulation and its related resisting response (Hartwell et al., 2011), meaning the ability and willingness to resist the urge to use a substance. It has been hypothesised that this region is also responsible for desire regulation and the gratification deriving from pleasure (Hartwell et al., 2011).

Over the last few years, several studies have demonstrated that neurostimulation techniques such as transcranial direct current stimulation (tDCS), which target the dorsolateral prefrontal cortex, can reduce carving (Boggio et al., 2008) as well as its related dysfunctional behaviour (Rachid, 2016). Furthermore, such techniques appear to have long-term effects. This means that even if the treatment is normally concluded within a limited number of therapy sessions, it can still have long-term effects.

In particular, tDCS is a non-invasive and painless technique with only mild adverse effects, which, if they appear, are limited to a slight itchy sensation over the stimulation site. It is simple and easy to perform. It is a procedure that entails modulated brain excitability by placing electrodes over the scalp; these electrodes release a low-intensity current flow for a few minutes. It has also been shown that tDCS can modify cognitive processes by combining neural activity and impulsive behaviour (Fecteau et al., 2004).

As far as behaviour is concerned, several studies have proven the efficacy of cue-exposure therapy (CET) in reducing the stimulus reaction associated with the addictive substance and craving (Tiffany & Conklin, 2002).

For this study, the Pavlovian conditioning model has been used. Some contexts, situations and objects (e.g. a bottle, glass, the bar for an alcoholic) are repeatedly associated with the addictive substance (which is an unconditioned stimulus, US); the context, situations and objects are instead conditioned stimuli (CS). Consequently, these factors elicit an impulse to seek and take the substance (conditioned response, CR) as if it were an unconditioned response (UR). Due to this conditioned context, the addicted person feels the craving when they face a conditioned stimulus. Hence, the craving stimulus becomes a trigger for the dependent behaviour (Lee et al., 2007).

CET aims to completely erase the response associated with the stimulus connected to the addictive substance. In order to do this, it is necessary to repeatedly expose the dependent person to signals connected to the substance (i.e. conditioned response) that causes dependence (Lee et al., 2007); however, this is done by precluding consumption, which could otherwise be a unconditioned stimulus.

As a consequence, CET involves a conditioned response such as craving, physiological activation (e.g. heartbeat, skin conductance, etc.), attention and behaviour biases that are connected to seeking the substance and activated by stimuli that have been previously associated with that substance (Ferreri et al., 2018).

Research on the application of CET (using scripts, photographs, videos, and objects related to drug consumption) has also helped to better understand those situations that lead to continuous substance use, as well as those factors that produce relapse (Conklin and Tiffany, 2002).

In addition, many studies have focused on salient stimuli, meaning videos and images that are offered via tablets and computers to patients during a neurostimulation session. This approach has proven to significantly reduce the outcome connected to dependence among addicts (Li et al., 2020; Carl et al., 2020). It may be possible that combining training and cognitive-behavioural therapy with neural stimulation can boost the therapeutic effects; this approach can also result in significant improvement in maintaining abstinence and reducing craving.

In light of the above, the main objective of this study is to evaluate the efficacy of neural stimulation in subjects who agreed to undertake CET training. This CET training and neural stimulation is expected to reduce craving and impulsivity levels (Bonfiglio et al., 2020), thus also reducing the impulse to use substances and reinforcing coping mechanisms such as the ability to resist substance consumption. By doing so, it is hoped that this approach will prove that these two techniques have to be applied together in order to obtain significant behavioural and neural changes.

## 2 Methodology

### *2.1 Instruments*

*Barratt Impulsiveness Scale.* BIS-11 is one of the most commonly used tests to measure impulsiveness (Patton et al., 1995). It comprises 30 items, which can yield total impulsivity as well as three related subscales: (a) attentive impulsiveness; (b) motor impulsiveness and (c) non-planning impulsiveness (Fossati et al., 2001). It has a four-point scale (0-not at all, 4-a lot).

*Symptom Checklist-90.* SCL-90 is a self-reporting instrument including 90 items where the subjects are asked to report on whether or not they have experienced specific symptoms in the 15 days prior to taking the test (Derogatis & Savitz, 1999). It consists of 9 primary symptoms dimensions: somatisation, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism. It has a four-point scale (0-not at all, 4-a lot).

*Self-efficacy and desire scale (SAD).* SAD comprises 27 items that describe several situations. Each situation includes two sets of options presented in two columns on which the subjects have to respectively choose their craving for the substance and their perceived ability to resist its usage (Minervini et al., 2011). It is possible to generate the total score and three subtotals resulting from the three related subscales: positive emotions and social situations, negative emotions and potentially critical situations, habits and abstinence. It makes use of a 10-point scale (substance desire: × from 0 “minimal desire” to 10 “maximum desire”; resist substance use perceived ability: from 0 “minimal ability to resist” to 10 “maximum ability to resist”).

*ASI (Addiction Severity Index).* ASI is based on a semi-structured multidimensional interview that aims to rate the severity index of substance addiction. Its 55 questions were designed to establish the intensity and frequency of the problems connected to drug-use within the previous 30 days. Patients are also asked to provide a self-assessment of their physical and mental condition and their relationship with their family. In particular, ASI seeks to investigate these general areas: alcohol and drug abuse, emotional and physical health, employment, family relations and illegal activity. It is extensively employed and has been translated into more than 20 languages (McLellan et al., 1992).

## 2.2 Cue Eliciting Training

The training session involved presenting 30 stimulating visual prompts and 10 neutral visual prompts. The former had been previously agreed upon with the individual subject and selected from a database of images that recalled several addictive substances. The latter were the same for all subjects.

Each image appeared on the screen moving from left to right, right to left or toward the subject scrolling from the bottom up. Each image remained on screen for 5 s.

At the end of the training session, each subject was asked to relax for a while in order to allow them to decompress if viewing the stimulating visual prompts had caused them any physiological stress. During this phase, the subjects were shown a series of 20 relaxing photographs they had selected beforehand; these visual prompts did not recall any addictive substances and were in contrast with the stimulating visual prompts already reviewed. These relaxing prompts were shown twice for 5 s each, for a total of 40 sequences and were accompanied by background music or songs previously selected by the subject to help them to relax.

## 2.3 tDCS Neurostimulation

For this experiment, tDCS was applied by using BrainStim stimulation devices (EMS, Italy), with pairs of silicone-coated electrodes (35 cm$^2$) that were inserted into sponges soaked in saline solution for EEG.

The anode was placed on a stimulation site on the scalp corresponding to the left dorsolateral prefrontal cortex (F3 location in the EEG 10–20 international system). These brain placement sites were chosen because, according to the existing literature, the left prefrontal cortex is responsible for controlling craving, while the anterior cingulate cortex is responsible for impulsive reactions and craving control (Hayashi et al., 2013). A 2 mA current intensity was applied for 20 min.

The control group was subject to a sham tDCS wherein the electrodes were placed on the same sites on the scalp, but the current intensity of the stimulator was gradually reduced to zero after a 20-min treatment. By doing so, the subjects did not know which procedure they were undergoing, since the typical tactile sensation associated with tDCS was experienced only at the beginning of the stimulation process (Brunoni et al., 2014).

### 2.4 Subjects and Procedures

A total of 40 subjects were selected for this experiment and all were patients that had been hospitalised for their pathological dependence in a rehabilitation centre in Lombardy, Italy. The subjects were recruited on a voluntary basis and according to specific including and excluding criteria that were used during a preliminary interview. The including criteria were: (1) being 18 years old or older; (2) having been diagnosed with substance addiction according to the DSM 5; (3) stable clinical conditions; (4) having abstained from substances for at least 50 days. The excluding criteria were: (1) suffering from epilepsy; (2) displaying severe clinical symptoms connected to abstinence; (3) severe psychiatric comorbidity; (4) convulsions and delirium tremens during periods of abstinence; (5) being already involved in other training experiments or other neuromodulation treatments; (6) any other contraindication to non-invasive brain electric stimulation, e.g. patients with intracranial metallic implants. All subjects signed a form providing written consent to the processing of their personal data for research purposes. This research project was approved by the Ethics Committee of the Department of Brain and Behavioural Science at the University of Pavia.

The treatment programme comprised 10 sessions. Each subject underwent two treatment sessions every week. Each session lasted 20–30 min. Each session included: (a) tDCS (active or sham); (b) cue eliciting; (c) cue relaxing.

This trial was designed with three experimental conditions:

(a) condition 1: the subjects were trained using an active tDCS;
(b) condition 2: the subjects were trained using a sham tDCS;
(c) condition 3: the subjects did not have any treatment.

Self-evaluation questionnaires were administered before starting the treatment programme (T0), after finishing the training programme (T1) and 1 month after the end of the treatment (T2). The experimental group was given the questionnaires the day after their interview during which their demographic data were collected and the stimuli were agreed upon. This was a blind trial for the experimental group, but not for the research team involved in administering the trial.

Table 1 reports on the frequency distribution, the mean values and standard deviation for the collected demographic data and diagnoses under scrutiny.

According to the Symptoms Checklist (SCL), the data shows no significant differences across the three groups in terms of their symptoms (all $p > 0.05$). This

**Table 1** Frequency distribution, mean and standard deviation for demographic data

| Variables | | Subjects of condition 1 | Subjects of condition 2 | Subjects of condition 3 |
|---|---|---|---|---|
| | | $n = 10$ | $n = 10$ | $n = 20$ |
| Demographic data | Age | 33 | 41.6 | 40.5 |
| | | ± 11.6 | ± 16.6 | ± 9.1 |
| | Education | 10.9 | 11.7 | 11.5 |
| | | ± 3.4 | ± 4.5 | ± 3.5 |
| | Sex | 7 M | 6 M | 17 M |
| | | 3 F | 4 F | 3 F |
| Therapy | Pharmacological | 10 | 8 | 11 |
| | Substitutive | 5 | 3 | 1 |
| Clinical data | Poly-substance abusers | 4 | 7 | 12 |
| | First cocaine | 2 | 3 | 8 |
| | First alcohol | 4 | 4 | 4 |
| | First heroin | 4 | 3 | 8 |
| | Age first use primary substance | 22.7 | 28.1 | 22.2 |
| | | ± 8.1 | ± 17.3 | ± 8.7 |
| | Days of abstinence | 95.4 | 121.9 | 121.5 |
| | | ± 46.6 | ± 171.5 | ± 154.9 |

therefore shows that there was no difference between the three groups in psychiatric terms. Furthermore, no values above 1 were detected, which demonstrates that none of the subjects displayed severe psychiatric symptoms.

## 3 Data Analysis

The outcome measures were analysed using a Linear Generalised Model (ANOVA). For each outcome, the within-groups factor results from the three administration steps (Time: T0, T1 and T2; meaning pre, post e follow-up) and the between-groups factor results from the data obtained by analysing the three groups of subjects under the three conditions (Condition: condition 1, condition 2, condition 3). The mean-square error term was used to conduct Tukey's honestly significant difference (HSD) post-hoc tests to determine potential differences between conditions. Post-hoc tests were considered significant at $P \leq 0.05$, with Cohen's d effect sizes reported for all post-hoc comparisons.

## 4 Results

With regard to impulsiveness, a significant relation between condition vs. time (F4,58 = 11.4, $p < 0.001$, age-square = 0.99) was detected. Table 2 below presents the mean values of the "impulsiveness" variable under all three conditions (see Table 2).

As may be noticed, the level of impulsiveness tends to remain constant under all three conditions during T0, T1 and T2. The only exception is condition 2, where the level of impulsiveness increases during T1 and T2.

As for the desire to take the substance again, a significant effect in the relation condition vs. time was also detected (F4,58 = 2.49, $p < 0.05$, age-square = 0.67). Table 3 below includes the mean values regarding the "desire" variable under all three conditions (see Table 3).

The level of desire to use the substance tends to decrease between the pre- and post-treatment period under condition1. Conversely, it remains stable under

**Table 2** Mean and standard deviation of impulsiveness for the three condition at time T0, T1 and T2

| Time | Conditions | Means | Standard deviations | N |
|---|---|---|---|---|
| **T0** | Condition 1 | 68.89 | 12.966 | 9 |
| | Condition 2 | 71.10 | 8.212 | 10 |
| | Condition 3 | 65.54 | 11.370 | 13 |
| | Total | 68.22 | 10.901 | 32 |
| **T1** | Condition 1 | 69.22 | 14.158 | 9 |
| | Condition 2 | 71.10 | 8.212 | 10 |
| | Condition 3 | 65.54 | 11.370 | 13 |
| | Total | 68.31 | 11.284 | 32 |
| **T2** | Condition 1 | 67.33 | 14.586 | 9 |
| | Condition 2 | 81.20 | 10.293 | 10 |
| | Condition 3 | 62.85 | 7.777 | 13 |
| | Total | 69.84 | 13.154 | 32 |

**Table 3** Mean and standard deviation of desire value for the three condition at time T0, T1 and T2

| Time | Conditions | Means | Standard deviations | N |
|---|---|---|---|---|
| **T0** | Condition 1 | 171.11 | 26.521 | 9 |
| | Condition 2 | 178.97 | 31.152 | 10 |
| | Condition 3 | 171.97 | 19.906 | 13 |
| | Total | 173.92 | 25.075 | 32 |
| **T1** | Condition 1 | 154.56 | 21.431 | 9 |
| | Condition 2 | 178.97 | 31.152 | 10 |
| | Condition 3 | 171.97 | 19.906 | 13 |
| | Total | 169.26 | 25.494 | 32 |
| **T2** | Condition 1 | 165.27 | 36.830 | 9 |
| | Condition 2 | 180.67 | 21.810 | 10 |
| | Condition 3 | 189.54 | 9.588 | 13 |
| | Total | 179.94 | 25.001 | 32 |

**Table 4** Mean and standard deviation for ability to resist for the three condition at time T0, T1 and T2

| Time | Conditions | Means | Standard deviations | N |
|---|---|---|---|---|
| **T0** | Condition 1 | 171.86 | 20.24 | 9 |
| | Condition 2 | 178.48 | 34.95 | 10 |
| | Condition 3 | 178.25 | 23.93 | 13 |
| | Total | 176.53 | 26.28 | 32 |
| **T1** | Condition 1 | 176.20 | 16.85 | 9 |
| | Condition 2 | 178.48 | 34.95 | 10 |
| | Condition 3 | 178.25 | 23.93 | 13 |
| | Total | 177.74 | 25.51 | 32 |
| **T2** | Condition 1 | 176.06 | 33.65 | 9 |
| | Condition 2 | 184.18 | 20.20 | 10 |
| | Condition 3 | 146.85 | 11.39 | 13 |
| | Total | 166.73 | 27.39 | 32 |

**Table 5** Mean and standard deviation for severity of addiction for the three condition at time T0, T1 and T2

| Time | Conditions | Means | Standard deviations | N |
|---|---|---|---|---|
| **T0** | Condition 1 | 1.49 | .61 | 9 |
| | Condition 2 | 1.25 | .34 | 10 |
| | Condition 3 | 1.33 | .39 | 13 |
| | Total | 1.35 | .44 | 32 |
| **T1** | Condition 1 | .94 | .72 | 9 |
| | Condition 2 | 1.25 | .34 | 10 |
| | Condition 3 | 1.33 | .39 | 13 |
| | Total | 1.19 | .50 | 32 |
| **T2** | Condition 1 | .98 | .62 | 9 |
| | Condition 2 | .84 | .56 | 10 |
| | Condition 3 | 1.23 | .45 | 13 |
| | Total | 1.04 | .54 | 32 |

condition 2 and condition 3. The craving for the substance tends to increase under all three conditions.

As for the ability to resist substance-taking, a key role seems to be played by the condition vs. time interaction ($F_{4,58} = 6.30$, $p < 0.001$, age-square = 0.98). Table 4 below includes all the mean values of this variable under all three conditions (Table 4).

The ability to resist substance-taking tends to remain stable within the T0 and T1 period under all three conditions. Under condition 3, it decreases considerably between T1 and T2, while it remains stable under the other conditions.

As regards the severity of the subjects' addiction, a significant factor seems to be the condition vs. time interaction ($F_{4,58} = 4.75$, $p < 0.001$, age-square = 0.93). Table 5 below presents the mean values detected under all three conditions (see Table 5).

As may be noted, the level of severity of the subjects' dependence tends to decrease under condition 1 between T0 and T1, while it remains stable under the other two conditions. This variable instead tends to decrease under condition 1 and condition 2 between T1 and T2, while it remains stable under condition 3.

## 5 Discussion

This experimental project has returned results that in part confirm the research hypotheses laid out here. Firstly, it appears clear that the increase in the impulsiveness level between T1 and T2 among subjects under condition 2 does not tend to decrease. This may depend on the fact that subjects under this condition did not undertake a tDCS procedure that, as demonstrated, can reduce impulsiveness and craving.

It may be noted that the level of craving decreases for all the subjects under condition1 while it remains stable for those under conditions 1 and 3 between T0 and T1. Conversely, craving tends to increase between T1 and T2 for the subjects under conditions 2 and 3, meaning during the period when none of the subjects were undergoing treatment; nevertheless, it remains stable for the subjects under condition 1. The latter result confirms the hypothesis that tDCS does have an effect on reducing craving and consequently, the desire and impulsiveness linked to seeking and taking a substance. This effect is due to the subjects' neurostimulation and continues even after the treatment is concluded, thus proving its long-term efficacy. The subjects under conditions 2 and 3 did not undertake any treatment that was directly targeting craving reduction between T1 and T2. Even if these subjects' craving level between T0 and T1 remained stable, it increased between T1 and T2, probably due to prolonged abstinence from the substance.

It is interesting to note that the ability to resist craving remains stable under all three conditions, despite the fact that under conditions 2 and 3 craving tends to increase progressively between one stage and another. It seems therefore safe to suggest that the CET training might have had an effect on subjects' ability to resist and find coping strategies to avoid relapse, even if it did not help reduce craving. Consequently, in the long-term and due to prolonged abstinence, only the subjects in group 3 experienced a decrease in their ability to resist craving.

What is more, the results obtained through ASI testing have shown that the treatment used for this project was successful in the subjects under conditions 1 and 2. The ASI test measures several aspects connected to addiction but craving and resistance to substance-taking are only two factors that have an indirect impact on the severity index. ASI testing yields data resulting from personal interviews that assess a wider range of variables, including legal aspects. The resulting data can therefore be considered as a reliable outcome in terms of the subject's general dependence condition, but it says very little about outcomes for specific symptoms such as craving and impulsiveness. That said, the results obtained regarding the group under scrutiny confirm the hypothesis that tDCS treatment in conjunction with CET training can effectively impact the severity of subjects' dependence, reducing it and contributing to progressive symptom remission.

All in all, this study has aimed to verify the efficacy of a neurostimulator tDCS treatment, coupled with CET training on a group of patients with addiction problems. This hypothesis appears to be partially confirmed. Neurostimulation has proved to be effective in reducing craving levels and dependence among the subjects under condition 1; it also contributes to reducing impulsiveness (monitored via a specific procedure) and increasing the ability to resist substance-taking. The CET training, on the other hand, does not seem to help reduce craving and impulsiveness levels, thus defying our expectations and hypotheses. However, it seems to help control the urge to take a substance. This data refers to the subjects under condition 2 and seems to be confirmed for the treatment period between T0 and T1, but it could not be confirmed for the following period between T1 and T2.

Interestingly, the subjects under condition 3, who can be defined as the control group, seem to confirm our hypothesis regarding the possible outcome in the treatment period between T0 and T1; however, they did not confirm our hypothesis for the T1 to T2 treatment period, when these subjects display the same outcome obtained by group 2. This is likely to depend on the fact that the CET training added a partial effect that was limited to the treatment period, while the tDCS treatment had long-term effects.

This study clearly has some limitations, which need to be taken into consideration. Firstly, the subjects who partook in this experiment were hospitalised in a rehabilitation centre where contingent factors can be very difficult to control. In addition, the criteria for choosing these subjects were that they be involved in similar group or individual activities (e.g. psychotherapy meetings), undergoing similar treatments and more or less experiencing similar conditions (e.g. being allowed out of the centre the same number of times or receiving an equal number of family visits). Nevertheless, it was impossible to control all these variables throughout the treatment period; therefore, the fact that such variables may have indirectly influenced the treatment outcome cannot be discounted.

In addition, the gender percentage is significantly unbalanced with a much higher number of male participants; also, the selection process was not random, and a double-blind experimental procedure could not be undertaken due to organization issues within the rehabilitation centre. That said, all the subjects were part of a rehabilitation programme and helped by a team of operators that actively collaborated on a daily basis with the research team involved in this experiment.

It is proposed that these limits be overcome to the extent possible in future research and that another condition be added with subjects solely treated using tDCS. In addition, we aim to conduct another experiment with subjects that will be solely treated with tDCS without training.

## References

Addolorato, G., Leggio, L., Abenavoli, L., Agabio, R., Caputo, F., Capristo, E., Colombo, G., Gessa, G. L., & Gasbarrini, G. (2006). Baclofen in the treatment of alcohol withdrawal syndrome: A comparative study vs. diazepam. *The American Journal of Medicine, 119*(3), 276–e13.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience, 8*(11), 1458–1463.

Boggio, P. S., Sultani, N., Fecteau, S., Merabet, L., Mecca, T., Pascual-Leone, A., Basaglia, A., & Fregni, F. (2008). Prefrontal cortex modulation using transcranial DC stimulation reduces alcohol craving: A double-blind, sham-controlled study. *Drug and Alcohol Dependence, 92*(1–3), 55–60.

Bonfiglio, N. S., Parodi, D., Rollo, D., Renati, R., Pessa, E., & Penna, M. P. (2020, June). Use of training with BCI (brain computer Interface) in the management of impulsivity. In *2020 IEEE international symposium on medical measurements and applications (MeMeA)* (pp. 1–5). IEEE.

Bonfiglio, N. S., Renati, R., Agus, M., & Penna, M. P. (2019). Validation of a substance craving questionnaire (SCQ) in Italian population. *Addictive Behaviors Reports, 9*, 100172.

Brunoni, A. R., Boggio, P. S., De Raedt, R., Benseñor, I. M., Lotufo, P. A., Namur, V., et al. (2014). Cognitive control therapy and transcranial direct current stimulation for depression: A randomized, double-blinded, controlled trial. *Journal of Affective Disorders, 162*, 43–49.

Carl, E., Liskiewicz, A., Rivard, C., Alberico, R., Belal, A., Mahoney, M. C., Quisenberry, A. J., Bickel, W. K., & Sheffer, C. E. (2020). Dosing parameters for the effects of high-frequency transcranial magnetic stimulation on smoking cessation: Study protocol for a randomized factorial sham-controlled clinical trial. *BMC Psychology, 8*, 1–14.

Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction, 97*(2), 155–167.

De Mulder, I., & Dom, G. (2012). Disulfiram as a treatment for cocaine dependency. *Tijdschrift voor Psychiatrie, 54*(1), 51.

Derogatis, L. R., & Savitz, K. L. (1999). *The SCL-90-R, brief symptom inventory, and matching clinical rating scales*.

Fecteau, J. H., Bell, A. H., & Munoz, D. P. (2004). Neural correlates of the automatic and goal-driven biases in orienting spatial attention. *Journal of Neurophysiology, 92*(3), 1728–1737.

Ferreri, F., Bourla, A., Mouchabac, S., & Karila, L. (2018). E-Addictology: An overview of new technologies for assessing and intervening in addictive behaviors. *Frontiers in Psychiatry, 9*, 51. https://doi.org/10.3389/fpsyt.2018.00051

Fossati, A., Di Ceglie, A., Acquarini, E., & Barratt, E. S. (2001). Psychometric properties of an Italian version of the Barratt impulsiveness Scale-11 (BIS-11) in nonclinical subjects. *Journal of Clinical Psychology, 57*(6), 815–828.

George, O., & Koob, G. F. (2011). Craving, context and the cortex. *Nature Neuroscience, 14*(4), 409–410.

Hartwell, K. J., Johnson, K. A., Li, X., Myrick, H., LeMatty, T., George, M. S., & Brady, K. T. (2011). Neural correlates of craving and resisting craving for tobacco in nicotine dependent smokers. *Addiction Biology, 16*(4), 654–666.

Hayashi, T., Ko, J. H., Strafella, A. P., & Dagher, A. (2013). Dorsolateral prefrontal and orbitofrontal cortex interactions during self-control of cigarette craving. *Proceedings of the National Academy of Sciences, 110*(11), 4422–4427.

Heinz, A., Beck, A., Wrase, J., Mohr, J., Obermayer, K., Gallinat, J., & Puls, I. (2009). Neurotransmitter systems in alcohol dependence. *Pharmacopsychiatry, 42*(S 01), S95–S101.

Johnson, B. A. (2008). Update on neuropharmacological treatments for alcoholism: Scientific basis and clinical findings. *Biochemical Pharmacology, 75*(1), 34–56.

Kienast, T., Schlagenhauf, F., Rapp, M. A., Wrase, J., Daig, I., Buchholz, H. G., et al. (2013). Dopamine-modulated aversive emotion processing fails in alcohol-dependent patients. *Pharmacopsychiatry, 2*(04), 130–136.

Koob, G. F., & Le Moal, M. (2008). Neurobiological mechanisms for opponent motivational processes in addiction. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 363*(1507), 3113–3123.

Koob, G. F., & Volkow, N. D. (2010). Neurocircuitry of addiction. *Neuropsychopharmacology, 35*(1), 217–238.

Lee, J. H., Kwon, H., Choi, J., & Yang, B. H. (2007). Cue-exposure therapy to decrease alcohol craving in virtual environment. *Cyberpsychology & Behavior, 10*(5), 617–623.

Li, X., Hartwell, K. J., Henderson, S., Badran, B. W., Brady, K. T., & George, M. S. (2020). Two weeks of image-guided left dorsolateral prefrontal cortex repetitive transcranial magnetic stimulation improves smoking cessation: A double-blind, sham-controlled, Randomized Clinical Trial. *Brain Stimulation, 13*, 1271–1279.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., et al. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment, 9*(3), 199–213.

Minervini, I., Palandri, S., Bianchi, S., Bastiani, L., & Paffi, D. (2011). Desire and coping self-efficacy as craving measures in addiction: The self-efficacy and desire scale (SAD). *The Open Behavioral Science Journal, 5*(1), 1–7.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology, 51*(6), 7681–774.

Rachid, F. (2016). Neurostimulation techniques in the treatment of nicotine dependence: A review. *The American Journal on Addictions, 25*(6), 436–451.

Tiffany, S. T., Carter, B. L., & Singleton, E. G. (2000). Challenges in the manipulation, assessment and interpretation of craving relevant variables. *Addiction, 95*(8s2), 177–187.

Tiffany, S. T., & Conklin, C. A. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction, 97*, 155–167.

Wrase, J., Makris, N., Braus, D. F., Mann, K., Smolka, M. N., Kennedy, D. N., et al. (2008). Amygdala volume associated with alcohol abuse relapse and craving. *American Journal of Psychiatry, 165*(9), 1179–1184.

# Representing Behavior, Consciousness, Learning: Will a Purely Classical Artificial Intelligence Be Enough?

**Mario Rasetti**

**Abstract** The question of whether or not the most advanced methods of Artificial Intelligence may be sufficient to implement a faithful representation of Behavior, Consciousness, Learning at classical level, with no need to resort to quantum information techniques.

**Keywords** Artificial intelligence · Behavior · Behavioral economics · Constructor theory · Learning · Topological data analysis

## 1 Introduction: A Quest for Intelligence, Human Behavior, Machines That Learn

Understanding behavior is hard without a preliminary rigorous discussion of intelligence: this is a long story. Let's try to summarize it, guided by Peter Kugel (2004). It all starts in 1950, when Alan Turing famously wrote a paper (Turing, 1950) devoted to the relationship between, say, computing machinery and intelligence. He indeed already had a precise definition of 'computing machinery', in the form that today we refer to as the 'Turing machine' (TM) (Turing, 1936); what he lacked was a precise definition of intelligence. This is why, instead of discussing intelligence as an abstract notion (Kugel, 2002), he created what he called the imitation game.

If we try to adopt the imitation game as a definition of 'intelligence', we realize right away that it is not only imprecise, but incomplete and even misleading. Indeed, Turing's purpose in suggesting it, however, was not to give a definition of intelligence but, in his words "to draw a line as sharp as possible between what was relevant to understand intelligence and what was not"; namely he intended primarily to establish what intelligence is not, rather than what it is. In those times, this was

M. Rasetti (✉)
ISI Foundation, Torino, Italy

ISI Global Science Foundation, New York, NY, USA
e-mail: mario.rasetti@isi.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_10

apparently the best he could do; however, he made a few claims about what he thought intelligence might be and how it might be related to computing machinery. Particularly intriguing are: (1) that in his view the machinery of the computer should be powerful enough to produce intelligent behavior, with no new more powerful hardware being necessary.

Several years later Jack Copeland and Diane Proudfoot (1999) have claimed that what they define as 'hypercomputation' would instead be necessary; (2) that the machinery of the computer would have to be able to do much more than computation before it might be made to behave intelligently. Turing stated that intelligent behavior presumably implies a departure from the behavior necessary for computation, but not a severe one, that would not give rise to random behavior, or to repetitive loops. This sort of behavior has been later identified with super-recursive algorithms (Burgin, 1999); (3) the deviation required might imply allowing the computer to make mistakes (Turing, 1986), because "if a machine is expected to be infallible, it cannot also be intelligent".

Mathematical models of the mind in terms of which one could try to develop a characterization of intelligence that is more precise and, hopefully, more accurate than the imitation game were proposed in time, which pave the way to what we call today 'machine intelligence' (or artificial intelligence). Psychologists define intelligence roughly as the ability to 'acquire and apply knowledge'. This hints to the fact that potentially intelligent machines should have two basic components: one, the 'learner', able to acquire knowledge; the other, the 'doer' to transform in concrete applications the knowledge that the learner acquired. Assuming that knowledge can be somehow represented by computer programs, the learner can be thought of as a system that generates programs, and the doer as a system that runs the programs generated by the learner to do useful things. It is clearly not enough for a machine to be defined as intelligent merely to have these two components; however, it would qualify as an expert system.

The ability of learning is the ancestor of modern notion of 'machine learning'—on which we shall return extensively. But in order for a learner/doer system to be considered intelligent, those components will have to perform at a certain level and the crucial question remains open: what is this level? No doubt the level reached by much of today's Artificial Intelligence (AI) is not high enough. The learner contribution is still inadequate to the intelligence of the whole system. Intelligence requires a learner that can get along with instructions that provide fewer details than those required for a detailed description of the system's behavior: even in everyday life we tend to consider a person who has to be told exactly what to do as not very intelligent.

Humans appear to learn to do things from a variety of sources; from examples, from vague instructions, from analogies and the like, from examples. One can think of a learner-from-examples as a system which succeeds in developing programs that can simulate the behavior of devices whose inner workings it cannot examine. All it is allowed to work with is the **behavior** (inputs $\leftrightarrow$ outputs) of the device. Of course, there are situations in which people simply cannot be told what the algorithm is (when they begin learning their native language, there seems to be no language in terms of which they can be told) and situations in which learning from examples is easier. Even if children could learn from detailed instructions, very few parents could produce them. Most don't even try; they just point to a few examples and let the child's mind do the rest. They can do that, presumably, because the child is intelligent; because it can learn from examples.

A learner that tries to develop programs from examples can be thought of as a system that strives to solve a 'black box identification problem'; namely, given the behavior of a device inside of which (a black box) it cannot look, generate a program that exactly duplicates the system's behavior. The problem is not too difficult to tackle (at least in principle) if the behavior is finite and the learner has access to all of it; it becomes however very difficult or even unhandily hard if the behavior is (at least potentially) infinite, as it has seen only a finite part of that behavior, and this means that it must go beyond the information given in a way that Turing computations cannot do, because it implies to be able to turn the infinite behavior of the sample-device into a finite program.

It is in the of the Theory of Behavioral Economics (TBE) (Frank, 1991; Thaler, 1994; Diamond & Vartiainen, 2007) that the notion of behavior has become a crucial keystone. The reason is that in economics this elusive feature bears on all sorts of proxies: social networks (emergence of opinions, preferences, beliefs); personal digital census (education, ethnicity, age group, sexual orientation); lifestyle; health; use of e-commerce channels, etc. Thus, behavior has added a new boundary and created a novel challenge to (AI).

In an ideal definition of consumers' behavior, things such as failures in repaying debts, framing perspectives or price anchors should not have any bearing on choices, and decisions would be merely the result of a careful weighing of the deterministic balance of costs and benefits; informed exclusively by concrete, well-defined needs and preferences, with every decision rational. Herbert Simons' concept of 'bounded rationality' (Simon, 1955, 1957, 1969; Klaes & Sent, 2005) dismantles such ideal, bringing into play the notion that consumers' minds (just behavior) must be understood in the framework of the environment in which they evolved. Thus, decisions are not always and not necessarily optimal, and not only because human information processing has other severe restrictions, due both to incomplete information or knowledge and limits to computational as well as logical capacities. The 'behavioral' approach to Economics Theory postulates that people (consumers) are boundedly rational agents, with a limited ability to retrieve or elaborate information.

Exploring just how available information affects the quality and outcome of decisions, and what happens in situations where people avoid information altogether, Richard Thaler coined the concept of 'mental accounting' (Thaler, 1999): people think of value in relative rather than absolute terms, deriving their fulfillment not just from an object's value, but from the quality of the deal as well: its transaction utility. Consumers tend to work with the totality of their mental accounts: personal experience, information that they consider reliable, prompt response are the key factors that enable them to make good decisions; yet aware information avoidance takes place, circumstances in which people choose not to secure knowledge, even if freely available.

On the other hand, more and more, much of the decisions in our society are data-based; made either by humans with the assistance of machine intelligence or wholly by AI machines. It is reasonable to assume that AI may reduce the impact of bounded rationality, as AI processes reduce societal information asymmetry and improve decision-making, thus also driving economy to more calculative rationality. The open question is whether the intervention of AI in applications such as online trading and decision-making may change economic theories, having on them an impact bearing on issues such as rational choice and expectations, computational thinking, portfolio optimization, counterfactual reasoning.

One basic issue to understand is that behavioral economics is not alternative with respect to other ongoing models of economics; it concerns instead understanding how such models account for the realities of human decision making, and can adjust to them. There are several actions that behavioral economics should strive to be able to do: (1) help proving that presently available models of fully rational, selfish, utility maximizing behavior can do a pretty good job in predicting behavior only provided one endows the utility function with a number of non-standard 'behavioral' features, such as regret, envy, desires for fairness, reciprocity or conformity; (2) show that the limited reasoning capabilities of humans leads to fundamental biases that simply cannot be modeled within the standard framework of economics; (3) help resolve those ambiguities in economic theory that show up when there are multiple equilibria, so as to recover the theory predictive power. Behavioral economics is expected to be instrumental to find out the most likely equilibrium that may occur when inserting in the models observed and perhaps machine-assisted people's reasoning.

Behavioral economists also resort to 'process measures', namely methods providing quantitative hints about the cognitive and emotional processes underlying decision making. The limitations of verbal accounts of the causes of one's own behavior are clearly understandable, however AI may relatively easily provide 'process tracing' tools to assess what information economic agents use in making decisions (Sanfey et al., 2003) as well as about how the considered information is processed.

The most advanced (and difficult to perform on large scale) process measures used are brain scans, typically by functional Magnetic Resonance Imaging (fMRI) (Camerer & Loewenstein, 2003; Camerer et al., 2005). Though data collection on this is still in its infancy, this method allows scientists to determine which parts of an individual's brain are activated in response to a task or decision. The 'semantic map of the brain', recently elaborated (Huth et al., 2016), provides here an very promising interpretation tool. These imaging methods have been indeed applied to a diversity of economic tasks, including decision making under risk and uncertainty, inter-temporal choice, buying and selling behavior and strategic behavior in games. Of course, severe ambiguities still underlay the interpretation of neural data in such a complex context as behavioral economics, and a great deal of care must be exerted moving on these slippery grounds. However, the recent success of methods of

Topological Data Analysis (TDA) of fMRI data in defining behavioral equivalence classes in the brain action (Petri et al., 2014) make us confident that precious information may be gathered along these lines of thought.

The new issue is here is that we are faced with the challenge of acquiring intelligence of an economical system in which the agents are not simply human beings but human beings each equipped with one or more electronic prostheses, able to augment almost arbitrarily their mutual connectivity, to provide them communication channels and access to knowledge in a measure that has no precedents in history, to let them put to effect their decision or play out their role on scales (time and space), with a speed, with an extension that human beings alone up to a few years ago could not even imagine to reach. Once more ‘behavior’ is the key-word.

George Graham’s incipit to his contribution to the Stanford Encyclopedia of Philosophy devoted to behaviorism (Graham, 2019), reads “It has sometimes been said that ‘behave is what organisms do.’“ Yet, the notion of ‘machine behavior’, which extends to machines the capacity to behave, is by now a ‘piece of the furniture’ of the jargon of Machine Intelligence science (i.e., AI). These two naïve observations imply a chain of effects. One cannot fully understand human behavior if it is decoupled from the context in which it occurs, and similarly one cannot really talk about machine behavior without the integrated considerations of the algorithms it operates with and the social environment in which such algorithms operate.

## 2 Artificial Intelligence and Machine Learning

Whenever studying behavior, in whatever context it may be interesting to study it, one needs to take into account at the agent level, where behavior is typically meant to refer to the outcome of human-human interactions, other forms of interaction: machine with machine and, above all, augmented human (namely humans equipped with some ‘intelligent’ prosthesis) with augmented humans. And this both in the passive sense: augmented humans’ behavior as object of study in itself, and in the active sense: the analysis of any social dynamics generated by interacting augmented humans, such as in the framework of behavioral economics.

In a somewhat simplified way, we can ascribe behavior to the combined and integrated effect of several features, which interact in non-linear way and give therefore rise to ‘emergent’ effects (in the sense of complexity science): (1) perception—the property of being endowed with appropriate ‘sensors’ and to be able to elaborate the signals thus received; (2) learning—the capacity to ‘learn’, namely to acquire from the perceived signals understanding, knowledge, skills, values, preferences (one could add ‘behavior’, with a curious circular, self-referential twist!); (3) reasoning—the ability to rationally elaborate what has been learnt, generating coherent, optimal, interactive feedback, and build, through: (4) predictivity—the capacity to extract from such extended expertise future scenarios; and (5) operative decisions—be they conceptual (models, strategies, interpretations, comprehension) or practical (action).

We argue that Machine Learning and its various 'layered' declinations, Deep Learning, Deep Reinforcement Learning, etc., in its present formulation is not sufficient for the task of dealing with phenomena based on or implying the use of a coherent, rigorous notion of behavior. The reasons for this are manifold. Fueled as they are by increasing computer power and new algorithms, ML techniques have become powerful tools for finding patterns in data. On the other hand, since quantum systems are able to produce counter-intuitive patterns believed not to be efficiently producible by classical systems, one can reasonably envisage a future in which quantum computers will outperform classical machines in ML tasks. For this reason, in the past few years plenty of research in Quantum Computation (QC) was focused on finding whether or not QC can help to improve classical ML algorithms. Ideas range from running computationally costly algorithms or their subroutines efficiently on a quantum computer to translating stochastic methods characteristic of data science into the language of quantum theory. The discussion bears in particular on the potential of a potential theory of quantum learning. Preliminary supervised and unsupervised quantum ML algorithms have been proposed for cluster assignment and cluster finding which provide an exponential speed-up over classical algorithms.

The road to understand and represent behavior in a complete way goes deep through the concept of **mind**, and this is a 'no go' constraint in itself, especially as a number of human mental capacities or some of their characteristics—such as the limits to learning or self-consciousness—are believed to outstrip Turing-computability \footnote{The discussion about the possibility of computation 'beyond Turing' is quite extended: see, e.g., (Siegelmann, 1995; Cabessa & Siegelmann, 2012)}, in other words to be 'non Turing computable' (Copeland, 2002; Rescorla, 2020). The complexity here stems out of the difficulty of defining the 'states of mind', a problem that the theory of Integrated Information Theory (IIT) tries to deal with (Tononi et al., 2016).

Let's focus first on a set of specific directions along which essential progress can and must be done. The first, crucial, is turning ML into a true theory. A hard task, which implies to be able to get rid of the dependence of ML from both structure and content of the training set and from the ontology that characterizes its strategy-defining tools. In principle, any procedure ultimately based on data mining, if efficient, should allow us to get what we look for (information and eventually knowledge) from any given data set—namely, irrespective of what is the subject matter data refer to—and lead us to get from data inspection not only answers consistent with what data in the set represent but the very questions to be asked. Also, we should be able to estimate a priori the level of learnability of the reference data set and to disentangle, with the highest possible precision, correlation from causation.

From all of this a novel, well defined, difficult challenge emerges for AI: that of representing the dynamics and evolution of (economic but not only) processes in which the new agents take part, i.e., including (intelligent) machines in the landscape, namely accounting as well for machine behavior (Rahwan et al., 2019). This requires to understand and then describe, represent, and measure an unprecedented form of self-referential intelligence.

The question we want to address is: does AI have the tools necessary to face this difficult task? And if not, or not yet, what does science need to do? The scientific approach to AI so far has never been called to tackle the subtleties and the rigor required to successfully attack such multi-faceted, complex, hard (both algorithmically and in the sense of complexity science) scenario; however, ongoing research is doing breathtaking steps forward, though in a cultural landscape that is an apparently inextricable (but we argue that it is not) mixture of lights and shadows. In what follows we want to review the state of the art of both methods and instruments that AI relies on and describe the most promising perspectives of the most advanced sectors of machine intelligence and data science we may expect.

Machine Learning (ML) (Bishop, 2006) deals with problems that are difficult to address with traditional programming techniques, such as classify a document according to some given criterion (e.g., spam) with sentiment analysis, estimate the probability that a credit card transaction is fraudulent or recognize an object in some given image. Typically, the result is a weighted combination of a large number of parameters, each contributing to the solution to some (perhaps small) degree.

As for Information Theory (IT), the processes taking place when ML operates can be pretty well represented, as we shall discuss, in the frame of Constructor Theory (CT), as they do share a common physical frame: neural networks, be they natural (human), NN, or artificial, ANN. ANNs, brick circuits of AI machines, aim to mimic the human brain connectome based on the concept that one way to think about the rational brain is that it works by accreting smaller abstractions into larger ones.

The CT approach is based on the assumption that the nature and properties of, say, 'information' (that we shall assume as example metaphor) follow entirely from the laws of physics. One must first express in exact terms the 'regularities' required in the laws of physics for the process (or procedure, or theoretical frame—in our example, information—as informally conceived, with its characteristic properties (like interoperability), to be instantiated in physical objects. Such regularities, thus expressed, have the status of new, though possibly only conjectured laws (principles) of physics. The theory (in our example of information) consists of those (assumed) principles of physics that explain the regularities in physical systems associated with the considered subject. A theory of this sort constitutes as well a framework, in which structure and features of a broad range of theories could be understood, provided that they obey these principles. The point is that the conceptual scheme considered does resemble some of the entities that appear in laws of physics, without regard to the specific media in which it is instantiated. This is referred to as the substrate independence of the theory, which therefore can also be moved from one type of medium to another while retaining all its properties (interoperability property). Interoperability is what makes human capabilities such as language and structured knowledge possible, as well as the possibility of biological adaptations that use symbolic codes, such as the genetic code. The theory focus has a counterfactual character: an object in a particular physical state cannot be said to carry it unless it might have been in a different state.

All these properties involve abstraction whereby one entity is represented symbolically by another. In the example of information, this implies that its representation in this form is no longer abstract, for it only exists when it is physically instantiated. In other words, the laws governing information are laws of physics. The attempt to incorporate ML into physics at a fundamental level, is absolutely innovative. All previous approaches have regarded it as an a priori mathematical or logical process. The approach suggested here is just the opposite, namely that the nature and properties of ML follow entirely from the laws of physics. For example, CT theory of information rests on first understanding computation in constructor-theoretic terms. This implies expressing information in terms of computation, not vice versa as it is usually done. Note again that here computation is not taken as an a priori concept and one seeks necessary and sufficient conditions for a physical process to instantiate it. Moreover, the conjecture is adopted that objective regularities exist in nature: the principles of physics. These can be conveniently expressed in terms of ‘computations’ and the property called ‘information’. The intuitive concept of information is associated with that of copying. An information variable is a clonable computation variable. Note that the constructor-theoretic mode of explanation has allowed this to be expressed as an exact, intrinsic property of the substrate. This provides the purely CT notion of information emancipated from its dependence on physics.

ML is the branch of AI concerning the construction and study of systems that can ‘learn’ from data. Its core capacity is that of representing data instances and functions evaluated on these instances in such a way as to allow for the recognition and then construction of the method the system will perform with on different sets of data instances. Keynote is the algorithm’s ability to perform accurately on new, previously unseen examples after having trained on a well-known learning data set. In other words, the core goal of a learner machine is to generalize from its experience. The training results are probability distributions obtained from a reduced scale experience on the data set, while the learner’s task is to extract something more general, so as to produce useful predictions in new cases. One can say that ML focuses on the discovery of previously unknown properties of the dataset. It should be kept in mind, however, that current ML systems operate almost exclusively in a model-blind (i.e., purely statistical) mode, namely not even incorporating very general assumptions or models of reality, such as the capacity of reasoning about retrospection or the outcome of interventions based on causal inference. This entails severe theoretical limits on their power and performance as well as, above all, their possibility of achieving a level intelligence in some way closer to humans’. A major endeavor, which will require to scientists to focus their efforts towards the construction of a new, more general and far-reaching mathematical architecture (the natural tool appears to be category theory) to make of AI and ML true pieces of science.

An intriguing question, talking about ‘learner machines’ is its connection with human mind—on which we shall return when introducing a possible quantum implementation of Integrated Information Theory—with human mind. A crucial problem is that for a computer (be it classical or quantum) one needs an enormous data-base to catch the basic structure whereas, on the contrary, in a human mind the necessary reference data-set does not need to store all training combinations, because apparently human mind can act in parallel on several ‘training states’. We conjecture that this is most probably due to the role played by topological invariants, which—as we shall discuss in the sequel—allow us to classify efficiently data into equivalence classes that may include a variety of spatial and thought entities. In a broader sense, this technique aims to ‘learning something useful’ about the environment within which the system operates as well as about how the system itself works. How gathered information is processed leads to the development of algorithms reflecting how to process high dimensional data and to deal with uncertainty.

Judea Pearl (Pearl & Mackenzie, 2018) argued that AI has been handicapped by such an incomplete understanding of what intelligence really is, and he figured out how to do a crucial step forward, devising a scheme called ‘Bayesian network’. Indeed, prime challenge in AI research is to program machines to associate a potential cause to a set of observable conditions.

For a vision for how truly intelligent machines would think, a key step is to replace reasoning by association with causal reasoning. The language of conventional process algebra is, so to speak, symmetric: if $X$ tells us about $Y$, then $Y$ tells us about $X$: deterministic relationships. Writing in mathematical form a simple fact—for example, that the upcoming storm causes the barometer to go down, and not the other way around—requires resorting to a form of asymmetric language able to capture our understanding that if $X$ causes $Y$ that does not mean that $Y$ causes $X$. There is here a strong hint to ‘non-commutative algebra’, and since the latter is typical of quantum physics (and of course of quantum computation), to the need of adopting the tools of quantum information.

Language is one of the tools whereby human intelligence expresses itself. The most recent discoveries of computational linguistics have evidenced how the central question in cognitive science is whether natural language provides essential combinatorial operations that are shared between diverse domains of thought. fMRI experiments on the role of linguistic mechanisms in forging the hierarchical structures of algebra, have shown how processing of the syntax-like operations of algebra does not rely on the neural mechanisms of natural language. Conversely, therefore, processing the syntax of language elicits the known substrate of linguistic competence, whereas algebraic operations involve bilateral parietal brain regions implicated, e.g., in the representation of magnitude. Natural Language Processing (NLP) techniques, on the other hand are mostly based on a direct approach, aimed exclusively to finding qualitative rules, not the underlying algebras or their manifestations. More powerful tools are needed for prediction and diagnosis, knowing as we do that this is merely the tip of human intelligence. If we want machines to reason about interventions and introspection, we must invoke more than interventions and introspection: causal models.

If we want automatic systems operate as expert systems able to replace the professionals, obviously we cannot approach the question but probabilistically. For long time this was believed not to be viable, as for a task like this standard probability calculations require exponential space and exponential time; but Pearl's Bayesian networks require polynomial time and moreover they are also quite transparent. The tools developed so far along these lines can enable machines to reason with uncertainty. What we need to pursue is however an even more challenging task than reasoning with cause and effect. This is only the intermediate step on the road to equip machines with a model of its environment. If a machine does not have a 'model of reality', it cannot be expected to behave intelligently in that reality. Eventually, machines will have to postulate such models on their own, and verify and refine them based on empirical evidence. In other words, AI must go across the same evolution road that always happened to science. In this way machines may perhaps 1 day acquire some sort of free will; we simply have to understand how to program them and what we gain out of it. Humans have the 'sensation' of free will. Evolution has equipped us with this sensation; thus, evidently, it serves some desirable computational function, whose first evidence is that we are able to communicate with each other counterfactually. Shall we succeed in equipping a machine with some capacity similar to this?

Whilst not all ML techniques have a natural description in terms of probability theory (as we said, typically Bayesian, because the notion of 'causation' plays a crucial role), many do, as it is the case for the framework of Graphical Models (the entanglement between graph theory and probability theory) that has enabled the unique efficient understanding and transference of ideas from statistical physics; essentially the notion of correlations, statistics, and entropy (as lack of information) (Jakulin, 2005; Skotarczak et al., 2018).

The circular chain

$$\text{Behavior} \leftrightarrow \text{Mind} \leftrightarrow \text{Intelligence} \leftrightarrow \text{Behavior}$$

touches also on the subtle, delicate question of 'Turing computability': there is a mathematical obstruction to reaching complete Turing computability of intelligence; an obstruction which can be circumvented only assuming that human reasoning is fundamentally unsound. Or, more precisely, the inferential scheme that Turing's computation provides us with is not adequate to the task of describing human mind. The most compelling original argument for the existence of such an obstruction was proposed by John Lucas (1961) and successively by Roger Penrose (1989, 1994), reviving questions originally proposed by Gödel and Turing himself.

# 3 Topology of Data in the Frame of Constructor Theory

## 3.1 *Beyond Machine Learning*

ML (Bishop, 2006) is very heuristic, this implies that its processes, in particular the construction of ontologies, are very much operator-dependent and even though improvements were devised to make the learning process more effective, such as deep reinforcement, it has at the moment little or no way to approach the questions of self-referentiality (machine behavior) and causation. TDA (Carlsson, 2009; Edelsbrunner & Harer, 2010; Zomorodian, 2009), in particular its far reaching version known as Topological Data Field Theory (TDFT) (Rasetti & Merelli, 2016), which can be fully unsupervised and can help to find questions before looking for answers, appear more promising. It is true that TDA is pretty hard computationally: it can however be efficiently used in conjunction with deep reinforcement learning, with the advantage of having a more rational approach to provide to the analysis a hierarchy of values. TDA naturally may naturally lead to classifying behavior in equivalence classes.

Coupling TDA with DL, is but a first approximation of some still unknown architecture, simpler and more elegant, as Nature is. In its TDFT formulation TDA leads us to understand that not only topological invariants are relevant—which provide a natural partition of data sets in equivalence classes—but the theory has an intrinsic algebra (related with the gauge group) which represents the process in a very rigorous way, of which the deep network is nothing but a layered, deeper black box. The deep force approach of the deep network requires enormous elaboration times (typically exponential, possibly reduced—by a clever, difficult use of Babai's method (2016)—to pseudo-polynomial). This as opposed to TDA, that coupled with simpler algorithms like *k–nn*,[1] requires polynomial time.

It is interesting to notice that ML and TDA (in particular in view of its TDFT version) may be thought of as having a common "physical" reference frame, provided by their implementation in terms of (A)NNs.[2] This is what makes the construction

---

[1] In statistics, the k–nearest neighbors algorithm (*k–nn*) is a non-parametric method used for both classification and regression. In these two cases, the input consists of the $k$ ($k \in \mathbb{N}$) closest training examples in the feature space. The output depends on whether *k–nn* is used for the former or the latter: (1) in *k–nn* classification, the output is a class membership, meaning that an object is classified by a vote of its neighbors, and the object is assigned to the class most common among its $k$ nearest neighbors. If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor; (2) in *k–nn* regression, the output is the 'property value' for the object, namely the average of the values of $k$ nearest neighbors. *k–nn* is a type of instance-based learning, in which the function is only approximated locally and all computation is deferred to function evaluation. Since for classification this algorithm relies on distance, normalizing the training data can strongly improve its accuracy.

[2] ANNs have a long history, from the early 40s, through the 70s: (McCulloch & Pitts, 1943; Hebb, 1949; Farley & Clark, 1954; Kleene, 1956; Rosenblatt, 1958; Minsky & Papert, 1969; Werbos, 1975); to the successive uptakes: (Dominic et al., 1991; Zell et al., 1993); to the recent successes: (Hinton, 2010; Lapuschkin et al., 2019).

of a more general theoretical framework possible, simply embedding both in a common scheme of CT. CT represents the fundamental laws of anything based on a physical support in terms of “possible vs. impossible” transformations and of the reasons for (im)possibilities. What we envisage is a TDAML representation in terms of a unifying theory expressed in the language of CT. This would lead to the capacity of including the task to face the intrinsic limits that the AI approach could possibly exhibit, including ‘decidability of learnability’, in the scheme. This in turn would allow us to represent decision-making processes through what is learnable or not, and why.

With TDA, algebraic topology methods integrated within the idea of treating data sets as spaces whose key property is that they are not vector spaces but topological spaces, whose ‘shape’ is relevant, have progressively earned pivotal interest in data analytics. In TDA the foresight is that appropriate mathematical tools, grounded on the subtle and profound notion of ‘space of data’—geometrical representation of large data sets as spaces—may enable us to incorporate data in a geometrical setting (topological) that leads to identify and control the hidden information patterns in an ever more effective way. Thus, data science is endowed with the capacity of playing its role in the most efficient way along its fundamental process:

$$\text{Data} \rightarrow \text{Information} \rightarrow \text{Knowledge} \rightarrow \text{Practical} \rightarrow \text{Wisdom}.$$

It is worth here to recall a few features of these crucial steps: (1) DATA: the method/design of data is not neutral collection; (2) INFO: also information can be processed in non-unique ways, however resorting to the analysis of the shape of data space to extract information as a pattern of correlations (as TDA does) amends this limitation; (3) KNOWLEDGE: it is intended here as the set of correlation patterns of information patterns, namely ‘knoledge’ $\equiv$ ‘correlated information’, and as such candidates itself as the natural framework for action; (4) WISDOM: it comes from the collection of scenarios emerging from the mathematical (algorithmic) models of the system generated by its representation—virtual yet faithful knowledge, as reached at step (3)—and implies the possibility of making rational, evidence-based decisions accounting for different possible actions. Causation plays here a fundamental role.

TDA is a theoretical framework allowing for an extremely efficient exploration of large amounts of data because it provides an innovative data mining method based on a self-consistent, non-linear topological field theory of the space of data. It may greatly improve the efficiency of ML techniques in exploring data sets, as it does not need a training set but only a full knowledge of the data space topology. The approach is rooted in the inference of information from global rather than local data space features. It stems out of the integration of the deep mathematical tools of analysis of the data space provided by combinatorial (algebraic) topology, with those of formal language theory and theoretical computer science. TDA goes beyond the conventional complex networks theory because it replaces the notion of network, where all ‘interactions’ are ‘two-body’, with that of simplicial complex—a hypergraph with a very rich combinatorial structure, where interactions involve arbitrary numbers of vertices—which is what allows us to overcome efficiently the limitations of conventional data mining methods.

Why is topology the natural tool to handle large, high-dimensional, complex spaces of data? Because: (1) 'Qualitative information is what is relevant: data users aim to build knowledge, namely to understand how data is organized on large scale, hence global, though qualitative, information is what matters; and topology is the branch of mathematics that deals with qualitative rather than quantitative geometric information about a space (connectivity, classification of loops and higher dimensional manifolds, invariants, ...); (2) 'Metrics' are not theoretically justified: whereas in the physical sciences phenomena support self-contained theories which tell us exactly what metric to use, in life or social sciences this is not the case; topology, contrary to metric geometry, studies geometric properties in a way insensitive to metrics: it ignores distance function and replaces it with some measurable notion of 'connective nearness' (proximity); (3) 'Coordinates' are not natural: data is typically conveyed and received in the form of vector-like strings of symbols, yet the 'components' or linear combinations or norm of these 'pseudo-vectors' are not natural in any sense: the space of data is not a vector space. Properties of the data space depending on some choice of coordinates are therefore not relevant. Topology deals just with those properties of geometric objects that do not depend on coordinates but only on intrinsic geometric features. It is coordinate-free. And, last but certainly not least, (4) 'Summaries' are what has value: 'typical' (or 'characteristic') trends are what provides the information one is looking for; topology gives a global representation of the data space, as—through its inherent combinatorial structure—it captures at once the summary of all relevant features.

In other words, what topology does is to provide us with an explicit representation of the data space shape, irrespective of what the data are concerned with. This allows us to include the cases when one doesn't know what to look for, as the data space shape embeds (and hence is able to suggest) all plausible possible answers, and therefore the way to figure out which are the right questions to ask, consistent with those answers.

## *3.2 The Topology of Data*

As for the latter point, the conventional method of handling data is by a graph (indeed a 'network', say W) whose vertex set is the set of points of data space and two points are connected by an edge if and only if their 'proximity measure' (in the sense of Grothendieck topology (Artin, 1962)) is below some given value, say $\varepsilon$. On W one determines then the optimal choice of $\varepsilon$. Such an approach is, however, too local to extract in a reliable way (i.e., to be able to classify in equivalence classes) the hidden correlation patterns: in other words, it is not sufficient to obtain dependable, global summaries. Topology allows us instead to resort to the representation of the space of data by a new object, a $\Sigma$ simplicial complex $\Sigma$, say $\Sigma$. W is a graph designed to capture very well data local connectivity, however it ignores a wealth of higher order global features, which are instead well discerned by its natural completion $\Sigma$, the higher-dimensional object of which W is the 1-dimensional "skeleton". $\Sigma$ is a piece-wise-linear space built out of simple pieces (simplices) identified combinatorially along their faces. The interesting feature referred to above is that this accounts not only for two-body interactions (in a network between two nodes there is either one link or nothing), but for arbitrary $n$-body relations (higher dimensional simplices, of dimension up to $n$).

Deep reason of the success of TDA is that topological measures and observables are by construction very robust, and that moreover they permit to capture explicitly interactions between more than pairs of agents (the nodes), thus providing a framework to describe, quantify, and compare the 'global shape' of arbitrarily complex (data) spaces (Weaver, 1948). This is crucial because virtually all interesting complex systems can be thought of as living in either 'configuration' or 'phase' spaces, including those that can be approximately described using finite datasets, and can therefore be faithfully represented embedding the corresponding data set into a simplicial complex. The two main concepts used to achieve this are 'persistent homology' and 'topological simplification' (Battiston et al., 2020; Torres et al., 2020).

Persistent homology groups encode the shape of topological spaces by progressively finer and finer approximations, calling into play higher order analogs of links between nodes, in a relational structure able to describe explicitly interactions of more than two agents at the same time. Moreover, they allow us to identify noise vs. signal and reduce it. The process emphasizes those topological features in increasing dimensions (one-dimensional cycles, three-dimensional cavities, etc.) that survive through the sequence and hence characterize the shape of the dataset, letting us compare in a principled way arbitrary spaces with different 'measures': number of points, shape (invariants), etc. We can thus study the shape of correlation spaces between data space regions and how such shape changes. Functional, global, and localized homological information can all be used to track the system evolution in time, and thus fingerprint individual subjects.

Topological simplification (known as 'Mapper', from the name of the most famous algorithm which implements it) is a topological dimensionality reduction scheme, aimed to extract low-dimensional simplicial-complex backbones from high-dimensional datasets. We can then resort to the possibility of using this topological information to build a topological skeleton able to highlight dissimilarities both in structure and function of different behavioral pathways. Such structure can be further leveraged to build a 'topologically informed' map of feature spaces, thus improving and stream-lining the selection of features important for classifications (e.g., equivalence classes of correlation patterns).

It is crucial to notice that the topological descriptions of data evolution, namely the characterization of spaces in terms of their topological invariants as well as their variations in the presence of external interventions, are equally useful in understanding and modeling the ANNs representing them, and their capacity to learn (react to) new tasks. Topological approaches have been realized to allow NNs (artificial or natural, i.e., human) to take advantage of homological descriptors to better detect or craft adversarial attacks by exploiting the topology of learned manifolds, and to improve the interpretability of what actually happens inside the NNs as they learn to perform complex tasks.

The crossover between topology, neuroscience and artificial intelligence occurs as the capacities of neural networks, like those of the human connectome (Sporns, 2012), reside in how they represent data spaces internally. Just because brain functions are encoded in functional patterns, detecting which is a well-defined problem of comparison of spaces, topological invariants provide a common thread and a robust tool to understand both cognitive and behavioral processes and AI, in its physical implementation through neural nets.

## *3.3 Neural Networks*

A neural network (be it a natural NN, i.e. the brain cortex or a part thereof, or AN artificial ANN, namely a circuit designed to reproduce the functions of some natural net) is a graph $\mathcal{G}$—possibly, an hypergraph—represented as such by a triple $\{\mathfrak{N}, \mathfrak{A}, \varphi\}$, where $\mathfrak{N}$ is the set of vertices, the 'neurons'; $\mathfrak{A}$ a set of directed or undirected $n$-neuron connections $s$ over $\mathcal{G}$, $1 \leq n \leq N$, for some given $N$, $\mathfrak{A} = \{s | s \in \Sigma\ (\mathfrak{N})\}$, whose elementary constituting elements are the 'axons', pairs $(i,j)$, $i, j \in \mathfrak{N}$, while $s$ are the many-body 'connections' among neurons associated with the corresponding simplices $s$ of all allowed dimensions $n$ up to $N$, in $\Sigma = \Sigma\ (\mathfrak{N})$; φ is a function over $\mathfrak{A}$, $\varphi : \mathfrak{A} \mapsto \mathbb{R}$ that attributes 'weights' $w_s$ to all connections $s$. $w_s$ is zero for connections that do not exist in the network. Typically, a single neuron $j$ is connected to a (large) number of other neurons, through a set of gates, the 'synapses'. The tensor $\mathcal{W}$ collecting all elements $w_s$, properly distributed among the simplices $s$ of $\Sigma$, generalization of the Hinton diagram adopted for conventional network W, describes all allowed interactions among neurons. $\mathcal{W}$ can be designed as a circuit in ANNs, in such a way as to be able to represent any behavior of the network $\mathcal{G}$. In some way, giving $\mathcal{W}$ is equivalent to providing the equations of motion in a physical system; it generates the response of the network to any given input. Notice that if is a hypergraph, then the neural interactions are no longer exclusively two-body but in general many-body, as TDA permits. Complexity of thought, in this view, is measured by the range of smaller abstractions one can draw on (reduction of $\mathcal{W}$ to lower-dimensional blocks), as well as by the number of times one can combine lower-level abstractions into higher-level abstractions.

As already mentioned, ML is rather heuristic. Suppose we have a set of input-output pairs (the training set) : the problem of ML consists in guessing (and validating) first the map $\mathcal{T} : \mathfrak{in} \mapsto \mathfrak{out}$, and then implementing a procedure that leads to describing such problem's guessed solution with a model $\mathfrak{M}$. Typically, $\mathfrak{M}$ depends (or, to be more precise, is assumed to depend) on a set of parameters, $\Theta$ (i.e., one chooses a parametric class of functions). Further steps are the definition of a 'loss function' to compare the results of the model with the experimental values and the 'optimization', namely the individuation of the parameters in $\Theta$

which reduce the loss to minimum. Thus, Machine Learning problems are in fact optimization problems. Why then do we talk about learning? The point is that the solution to the optimization problem is not given in an analytical form; indeed, often there exists no closed form solution. For this reason, one has to resort to iterative techniques, typically 'gradient descent', to approximate the result progressively. It is this form of iteration over data that is understood as a way of progressive learning of the objective function based on the experience of past observations. The important feature here is that physical circuits (real pieces of the connectome or electronic circuits) can reproduce, to almost any desired level of approximation, neural nets—generalized or not—and hence the processes of ML that these can perform.

## 4 Constructor Theory Meets Learnability

The other crucial ingredient necessary to pursue progress along this ambitious, hard pathway comes from an apparently extraneous front: Constructor Theory. Modern Constructor Theory (CT) is a quite visionary extension of the John von Neumann's far-reaching idea of 'universal constructor'[3] (Burks, 1971), a self-replicating machine in a cellular automata environment, designed in the '40s, and conceived with no notion yet of a computer.

Revived and fully (and rigorously) reformulated by David Deutsch (2013), Deutsch & Marletto (2015), and Chiara Marletto (2016), CT was recently used to construct IT completely and solely in terms of which transformations of the ground physical systems may occur and which may not. This is indeed, in a nutshell, what constructor theory does in general: the basic principle of CT being that "All subsidiary theories are expressible entirely in terms of 'statements' about which physical transformations are possible and which are impossible, and why". The point is that CT regards science—even IT—not merely as an enterprise for the purpose of making predictions, but as an endeavor for discovering what the world is really like, how it behaves the way it does and why.

Like for information, nature and properties of ML follow entirely from the laws of physics. As in the CT theory of information, we expect that ML consists of proposed principles of physics that explain the regularities in physical systems that are associated with it. In other words, like information resembles entities that appear in laws of physics for a certain 'substrate', and this gives it a counter-factual character (an object in a particular physical state cannot be said to carry information unless it has been in a different state), analogously ML can be formulated on an ANN circuit, and this provides it of a substrate, and a counterfactual character. This representation of ML involves of course abstractions in that one entity is represented symbolically by another. Thus, the laws governing ML can be thought

---

[3] The details of the theory were published in von Neumann's book (von Neumann, 1966), completed by Arthur Burks after von Neumann's death.

of as laws of physics. This will allow us to express it in terms of computation. This is the key to providing the base for the recursive definition of ‘learning’ that we described (notice the strict analogy with the concept information, which is associated with that of copying). That’s why the representation of TDA∪ML in terms of a unifying framework expressed in the language of CT may lead to be able to deal with the hardest intrinsic limits that the AI approach could possibly exhibit, such as decidability of learnability. The latter refers to the possibility of representing decision-making processes through what is learnable or not and why; a crucial property related also to the fact that CT embodies naturally the feature of being able to implement an autopoietic (etymologically: self-producing) behavior (Maturana & Varela, 1980, 1987). This means to operate as a network of component-producing processes, with the property that the interaction among such components generates the very same sort of network of processes that produced them, yet constituting it as a distinct (autonomous) entity in the space in which it exists.

A decisive issue to complete the framework discussed is that of Learnability in ML. A sound mathematical foundation for ML through CT (once again the natural mathematical tool here is category theory (Spivak, 2014) quite general, powerful and far-reaching; combined, of course, with the more conventional formal logic and the theory of formal languages) will progressively improve our understanding and provide us with novel principles and frameworks to design new learning paradigms; in particular no go theorems. This bears on the fact that also ML cannot escape the curse that any advancement in mathematics must have a cost. In 1931, Kurt Gödel (1931) showed that in any system of axioms that is expressive enough to model arithmetic, the truth or falsehood of certain statements is not provable (it is undecidable). In the successive applications of this taxing notion, it was shown (Cohen, 1963) that Georg Cantor’s Continuum Hypothesis (CH)—which states that no set of distinct objects has a size (cardinality) larger than that of the integers but smaller than that of the real numbers—cannot be proved nor refuted using the standard axioms of mathematics.

It is now known that also ML does not escape the fate of Gödel’s incompleteness theorems. Recently Shai Ben-David et al. (2019), resorting to the formal equivalence between machine learnability and data compression (the process of encoding information using fewer bits than the original representation) were able to show that the solution to the respective optimization problems may be isomorphic to the proof of CH, and thus succeeded in constructing scenarios proving that learnability in ML may be undecidable in the sense of Gödel.

Of course, identifying the learnable is (it must be) a fundamental goal of ML: but to achieve it, one needs a robust mathematical framework, supporting the formal treatment of learnability. Conventional paradigms of ML fail to do this, as learnability cannot always be decided by standard axioms of mathematics, which, for example, are unable to provide any dimension-like quantity characterizing learnability in full generality. We argue that redefining such paradigms within the boundaries, rules and constraints of CT∪TDA may lead as well to defining efficiently such quantity.

# 5 Is ‘Quantum’ Necessary? The Integrated Information Theory Issue

## *5.1 Quantum Information and Consciousness*

In the quantum domain, the question of whether or not a quantum algorithm could successfully deal with classically non-computable functions has been long debated and remains essentially controversial. It has been shown that some problems, such as the halting problem for Turing machines, can be in some way successfully approached in a quantum perspective, but it is in fact widely believed that quantum computation cannot say anything general about computability. Contrary to this, quantum strategies have been suggested that define computability and its limits through a mixture of mathematical and physical principles, and in this perspective it was shown that quantum computation may be able to compute the non-computable provided suitable Hamiltonians could be constructed whose ground state represented the solution to classically incomputable problems. For example, a quantum algorithm for the classically non computable Hilbert’s tenth problem—whose challenge is to construct a general algorithm able to decide whether any given Diophantine equation (a polynomial equation with integer coefficients and a finite number of unknowns) has a solution with all unknowns taking integer values—the various formulations of which can be made isomorphic to the halting problem for Turing machines. Turing halting problem is known to be mathematically non computable, yet if quantum continuous variables and quantum adiabatic evolution could be physically implemented, classical computability constraints as posed by the Church-Turing thesis could be overcome.

Talking about human behavior, implies being able to say something viable about consciousness, which touches, as we shall see, on computability. The conceptual scheme that best fits the scheme we designed—including the possibility of reformulating it in terms of CT—is no doubt Tononi’s IIT (Tononi et al., 2016). Over the last decade IIT has emerged as one of the relevant tools in computational neuroscience of human mind. IIT aims indeed at providing a mechanistic, mathematically well-defined description of the neural correlates of consciousness. Its basic idea is to identify consciousness with the amount, suitably quantified, of cause/effect generation power in the neural network. The latter is assumed to be holistic, in the sense that it goes beyond and above the sum of its parts, which interact non-linearly. In order to implement this feature, one needs to quantify how certain (arbitrary) parts of the network, referred to as ‘mechanisms’, in a given state, irreducibly influence the future and constrain the past of other (arbitrary parts called (‘purviews’). Irreducibly means that the process cannot be represented (‘reduced’) as the separate and independent actions of parts of the mechanism over parts of the purview. Iterated at the global network level such process gives rise to a conceptual structure, which comprises a family of mechanisms and purviews, the latter representing the integrated core causes/effects of the former (Tononi, 2004, 2015; Tononi et al., 2015; Oizumi et al., 2014; Albantakis & Tononi, 2015).

What one looks for is a viable measure of the distance between this conceptual structure with the closest one obtainable from the set of all network partitions in all ways needed to quantitatively determine how much of the cause/effect dynamical network structure fails to be reducible to the sum of its parts. The minimal distance obtained in this way is, by definition, the Integrated Information (denoted by $\Phi$) of the network. IIT boldly postulates that the larger $\Phi$, the higher is the 'degree of consciousness' of the network in a particular given state. It should be remarked that the causal information-processing structure irreducibility of the network measured by is independent of the specific implementation of the brain circuitry; be it the connectome (real neurons, synapses and axons) or suitably complex ANN circuits. Quantum extensions of IIT have been proposed (Tegmark, 2015, 2016), however they refer to more abstract and general systems.

It is in view of the (algorithmic) complexity of the processes represented by IIT that the necessity emerges of paying special attention to its possible reformulation in a general and consistent version for interacting networks of finite-dimensional quantum systems. This because only the computational efficiency made available by the features of quantum information can make the conceptual model we are talking about—assuming the computability question is overcome—viable for a manageable theory. Most promising in the quantum information-theoretic framework is the approach due to Paolo Zanardi et al. (2018). It considers as neural system a network $\Lambda$ of qu-$d$its, the probability distributions being represented by non-commutative density matrices $\varrho$, constrained by Bayes' rules, and the related Markov processes identified with trace preserving completely positive maps $\mathcal{U}$. In this approach the irreducible cause/effect structure of the global network of IIT turns out to be encoded into a 'conceptual structure operator', quantum version of the classical counterpart. The minimal distance in norm of such operator from those obtained by all factorized versions of the network defines the quantum Integrated Information $\hat{\Phi}_q$.

## 5.2 *The Quantum Approach: Necessary But Not Sufficient*

Here is a concise outline of the idea. Within the set of all mechanisms in $\Lambda$, consider pairs $\mathcal{M}, \mathcal{P} \in \Lambda$ (where the mechanism set $\mathcal{P}$ is referred to as the 'purview' of $\mathscr{M}$; we denote by $\check{\mathcal{M}}$, $\check{\mathcal{P}}$ the complements of $\mathscr{M}$ and $\mathcal{P}$ in $\Lambda$, respectively). As mentioned, the network dynamical evolution is identified with the action of a (complete, positive) unital map $\mathcal{U}$. One wants to measure how the state—i.e., the probability distribution described quantum mechanically by a density matrix $\varrho$—of $\mathscr{M}$ at time $t$ (while $\check{\mathcal{M}}$ is in a maximally random state) conditions the state of $\mathcal{P}$ at time $t \pm 1$ (with also $\mathcal{P}$ maximally randomized). For any $\mathscr{M}$ we call a concept the triple $\{\varrho^{(c)}, \varrho^{(e)}, \varphi(\mathscr{M})\}$, where we denote by $\varrho^{(c)}$, $\varrho^{(e)}$, the density matrices (probability distributions) associated, respectively, with cause and effect, whereas $\varphi(\mathscr{M})$ is the maximum over all possible $\mathcal{P}$ 's of the effect ($e$) and cause ($c$), in the repertoire of $\mathscr{M}$, over the purview $\mathcal{P}$, formally (and manifestly, in quantum mechanical language) defined, for $x = c, e$, by:

$$\varrho^{(x)}\left(\mathcal{P}|\mathcal{M}\right) \doteq \mathrm{Tr}_{\check{\mathcal{P}}}\mathcal{U}^{(x)}\left(\Psi_{\mathcal{M}} \otimes \frac{\mathrm{I}_{\check{\mathcal{M}}}}{d^{\left|\check{\mathcal{M}}\right|}}\right); \Psi_{\mathcal{M}} \doteq \mathrm{Tr}_{\check{\mathcal{M}}}\Psi_{\Lambda},$$

where $\Psi_{\Lambda}$ is the network quantum state, while $\mathcal{U}^{(c)} \equiv \mathcal{U},\ \mathcal{U}^{(e)} = \mathcal{U}^{*}$ (the Hilbert-Schmidt dual of $\mathcal{U}$). The totality of concepts is our conceptual structure operator $\mathfrak{C}$. It is worth pointing out that the collection of triplescan be interpreted as the quantum version of the notion of the 'space of qualia' of the neurosciences. Given two conceptual structures $\mathfrak{C}_1, \mathfrak{C}_2$, we define the conceptual distance $\mathcal{D}\left(\mathfrak{C}_1, \mathfrak{C}_2\right) \doteq \frac{1}{2} \parallel \mathfrak{C}_1 - \mathfrak{C}_2 \parallel$. The measure of the Integrated Information of $\Lambda$ can finally be defined in this way: for $\mathfrak{P}$ the set of all $2^{|\Lambda|-1} - 1$ possible bi-partitions of $\Lambda$,

$$\Phi\left(\mathcal{U}\right) = \min_{\mathfrak{P}} \mathcal{D}\left(\mathfrak{C}\left(\mathcal{U}\right), \mathfrak{C}\left(\mathcal{U}_{\mathfrak{P}}\right)\right),$$

where, for $\mathfrak{P}\left(\Lambda\right) = \{\Lambda_1, \Lambda_2\}, \mathcal{U}_{\mathfrak{P}} = \mathcal{U}_{\Lambda_1} \otimes \mathcal{U}_{\Lambda_2}$.

As the algorithmic complexity of the computation of $\Phi$ is forbidding, the conclusion appears really to be that 'quantum' is necessary but possibly not sufficient to deal with the wide class of behaviors, in which the components of a complex system of concurrently executing agents interact with each other to achieve some global effect. Indeed, on a "macro" scale, this description fits well today's distributed transactions across the Internet; but on a "micro" scale, it is equally true for how functional computations are ultimately realized. All computations can be resolved into the interactions of large numbers of very simple agents: the complex behavior of the overall system is an "emergent property" of these interactions. Thus, the models of computation needed, which take interaction as their basic ingredient, yet combine sufficient expressive power to yield faithful descriptions of information, with sufficient mathematical structure and tractability to provide a basis for the formal analysis easily face the obstructions of excessive algorithmic complexity. The real question is computability, which can be unreachable because of the lack of learnability.

## 6 Conclusions

The approach presented tries to figure out the answerability in parallel of a number of hard questions: humans augmented by electronic prostheses (AI, for short), considered in the frame of TBE are quite different from the agents that economical sciences are used to deal with. It is this that implies a new strain on the analysis of their behavior and that requires better defined and farther reaching tools, enlarging as it does the domain these need to provide answers to. Indeed, besides human/human we have machine/machine interactions and human/machine interactions, namely augmented phase spaces of cognitive functions, more complex classification of behaviors, a novel role of mind (in the deep sense of self-consciousness) and of the living vs. inert quality of the physical support. For sure an unprecedented challenge for data science.

**Acknowledgments** The author acknowledges fruitful discussions and exchange of ideas with Riccardo Prodam.

*Dedication*: this paper is dedicated to the fond memory of Eliano Pessa; he would have enjoyed discussing it.

## References

Albantakis, L., & Tononi, G. (2015). The intrinsic cause-effect power of discrete dynamical systems; from elementary cellular automata to adapting Animats. *Entropy, 17*, 5472–5502.

Artin, M. (1962). *Grothendieck topologies*. Harvard University Press.

Babai L. (2016) *Graph Isomorphism in Quasipolynomial Time*. arXiv:1512.03547v2 [cs.DS].

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., & Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports, 874*, 1–92.

Ben-David, S., Hrubeš, P., Moran, S., Shpilka, A., & Yehudayoff, A. (2019). Learnability can be undecidable. *Nature Machine Intelligence, 44*(1), 44–48.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Burgin, M. S. (1999). Super-recursive algorithms as a tool for high performance computing. In *Proceedings of the high performance computing symposium 1999* (pp. 224–2228). UCSD Press.

Burks, A. W. (1971). *Essays on cellular automata*. University of Illinois Press.

Cabessa, J., & Siegelmann, H. T. (2012). The computational power of interactive recurrent neural networks. *Neural Computation, 24*(4), 996–1019.

Camerer, C., & Loewenstein, G. (2003). Behavioral economics: Past, present, future. In C. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in behavioral economics* (pp. 3–51). Russell Sage Foundation Press; Princeton University Press.

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature, 43*, 9–64.

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society, 46*(2), 255–308.

Cohen, P. J. (1963). The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 50*(6), 1143–1148.

Copeland, B. J. (2002). Hypercomputation. *Minds and Machines, 12*, 461–502.

Copeland, B. J., & Proudfoot, D. (1999). Alan Turing's forgotten ideas in computer science. *Scientific American, 280*, 76–81.

Deutsch, D. (2013). The philosophy of constructor theory. *Synthese, 190*(18), 4331–4359.

Deutsch, D., & Marletto, C. (2015). The constructor theory of information. *Proceedings of the Royal Society A, 471*, e20140540.

Diamond, P., & Vartiainen, H. (2007). *Behavioral economics and its applications*. Princeton University Press.

Dominic, S., Das, R., Whitley, D., & Anderson, C. (1991). Genetic reinforcement learning for neural networks. In *IJCNN-91- IEEE Seattle International Joint Conference on Neural Networks* (pp. 71–76). IEEE.

Edelsbrunner, H., & Harer, J. (2010). *Computational topology, an introduction*. American Mathematical Society.

Farley, B., & Clark, W. (1954). Simulation of self-organizing systems by digital computer. *Transactions of the IRE Professional Group on Information Theory, 4*(4), 76–84.

Frank, R. H. (1991). *Passions within reason: The strategic role of the emotions*. W. W. Norton.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathe matik und Physik, 38*(1), 173–198.

Graham, G. (2019). Behaviorism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University.

Hebb, D. (1949). *The organization of behavior*. Wiley.

Hinton, G. E. (2010). A practical guide to training restricted Boltzmann Machines. In *Mississauga Library Technical Report 2010–003*. University of Toronto.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453–458.

Jakulin, A. (2005), *Machine learning based on attribute interactions* (PhD thesis). University of Ljubljana.

Klaes, M., & Sent, E.-M. (2005). A conceptual history of the emergence of bounded rationality. *History of Political Economy, 37*(1), 27–59.

Kleene, S. C. (1956). Representation of events in nerve nets and finite automata. *Annals of Mathematics Studies, 34*, 3–41.

Kugel, P. (2002). Computing machines can't be intelligent (... and Turing said so). *Minds and Machines, 12*(4), 563–579.

Kugel, P. (2004). Towards a theory of intelligence. *Theoretical Computer Science, 317*, 13–30.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications, 10*, 1096.

Lucas, J. R. (1961). Minds machines and Gödel. *Philosophy, 36*, 112–127.

Marletto, C. (2016). The constructor theory of probability. *Proceedings of the Royal Society A, 472*, e20150883.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition*. Reidel.

Maturana, H. R., & Varela, F. J. (1987). *The Tree of Knowledge*. Shambhala.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*(4), 115–133.

Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology, 10*(5), e1003588.

Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The new science of cause and effect*. Basic Books.

Penrose, R. (1989). *The emperor's new mind*. Oxford University Press.

Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.

Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P. J., & Vaccarino, F. (2014). Homological scaffolds of brain functional networks. *Journal of the Royal Society Interface, 11*, 20140873.

Rahwan, I., et al. (2019). Machine behavior. *Nature, 568*, 477–486.

Rasetti, M., & Merelli, E. (2016). Topological field theory of data: Mining data beyond complex networks. In P. Contucci & C. Giardinà (Eds.), *Advances in disodered systems, random processes and some applications* (pp. 1–42). Cambridge University Press.

Rescorla, M. (2020). The computational theory of mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755–1758.

Siegelmann, H. T. (1995). Computation beyond the turing limit. *Science, 268*(5210), 545–548.

Simon, H. H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*(1), 99–118.

Simon, H. H. (1957). *Models of man*. Wiley.

Simon, H. H. (1969). *The sciences of the artificial*. MIT Press.

Skotarczak, E., Dobek, A., & Moliński, K. (2018). Entropy as a measure of dependency for categorized data. *Biometrical Letters, 55*(2), 233–243.

Spivak, D. I. (2014). *Category theory for the sciences*. IT Press.

Sporns, O. (2012). *Discovering the human connectome*. MIT Press.

Tegmark, M. (2015). Consciousness as a state of matter. *Chaos, Solitons & Fractals, 76*, 238–270.

Tegmark, M. (2016). Improved measures of integrated information. *PLoS Computational Biology, 12*(11), e1005123.

Thaler, R. H. (1994). *The Winner's Curse: Paradoxes and anomalies of economic life*. Princeton University Press.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*, 183–206.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5*, 42.

Tononi, G. (2015). Integrated information theory. *Scholarpedia, 10*(1), 4164.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience, 17*(7), 450–461.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B, 370*, e20140167.

Torres, L., Bassett, D. S., Blevins, A. S., & Eliassi-Rad, T. (2020). *The why, how, and when of representations for complex systems*. arXiv: 2006.02870v1 [cs.SI].

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 2–42*(1), 230–265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

Turing, A. M. (1986). Lecture to The London Mathematical Society, 20 February1947. In B. E. Carpenter & R. N. Doran (Eds.), *A.M. Turing's ACE Report and Other Papers*. MIT Press.

von Neumann, J. (1966). *Theory of self-reproducing automata*. University of Illinois Press.

Weaver, W. (1948). Science and complexity. *American Scientist, 36*, 536–546.

Werbos, P. J. (1975). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Harvard University Press.

Zanardi, P., Tomka, M., & Campos-Venuti, L. (2018). *Towards Quantum Integrated Information Theory*. arXiv: 1806.01421v2 [quant ph].

Zell, A., Mache, N., Hüttel, M., & Vogt, M. (1993). Simulation Neuronaler Netze auf Massiv Parallelen Rechnern. In H. Reichel (Ed.), *Informatik—Wirtschaft—Gesellschaft* (pp. 495–502). Springer.

Zomorodian, A. J. (2009). *Topology of computing*. Cambridge University Press.

# On Randomness and Origin of Life

**Roberto Serra**

**Abstract** It is argued that, in order to understand how life might have been originated under abiotic conditions, it is necessary to describe the appearance of both primitive cells and self-replicating sets of molecules. It is observed that, under quite general assumptions, the fluctuations in the internal composition of small protocells may be an important factor, which allows self-replication to take place by limiting the number of different types of molecular species and of reactions. The notion of a shadow biosphere is also briefly discussed.

**Keywords** Binary polymer · Diffusion · Membrane · Protocell · Random fluctuation · Replication · Reproduction · Shadow biosphere · Vesicle

## 1 Introduction

When Gianfranco Minati asked me to contribute a paper to honour the memory of Eliano Pessa, I readily accepted. I really liked Eliano as a person and as a scientist, and I admired his efforts to build an Italian community in what might be called "non-orhodox systems science". Eliano was a daring and non-conventional scientist, two features worth of great praise.

After much thinking, and a couple of false starts, I decided to describe and to comment a property which my colleague and friend Marco Villani and I observed in our recent studies of protocells, and which may be related to some deep aspects of the emergence of life in an abiotic environment.

---

R. Serra (✉)
Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, Modena, Italy

European Centre for Living Technology, Venice, Italy

Institute of Advanced Studies, Amsterdam, The Netherlands
e-mail: rserra@unimore.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_11

I have been working on several kinds of complex systems for a long time, and in 2004 I started to develop models of the dynamics of protocells, i.e. cell-like structures which are much simpler than biological cells, yet share some of their key properties, including growth and reproduction. There is some confusion in the popular press, but also in the scientific literature, since the term *protocell* is sometimes used to denote any kind of "simplified" cell, including those which are obtained starting from material of biological origin (e.g. by removing part of the genes of a bacterium). Here I will consider only those protocells that are synthesized, either by spontaneous or engineered processes, using only abiotic material. Let me also remark that the synthesis of this kind of true protocells, able to go through a series of replication cycles, has not yet been achieved.

It is clear that protocells have close connections with the problem of the origin of life. However, one can also look at them as a promising new technology, which may be interesting for theoretical reasons as well as for possible applications (like synthesis of new drugs, removal of contaminants, etc). I have so far followed this prudential approach, avoiding to publish speculations about the problem of the origin of life. However, it is crystal clear that this is actually the main reason of scientific interest for doing research on protocells, so I will deal with it in this paper. I hope Eliano would have liked this choice.

Of course, I have no aim at completeness, but I will try to point out some aspects that seem to me really interesting. There are very many scientific papers about protocells, and even more about the origin of life. I will avoid filling this paper with too many references, addressing the interested reader to those contained in a book I wrote a few years ago (Serra & Villani, 2017) and limiting to explicitly mention further references where strictly necessary.

Most books and papers deal with the problem of identifying possible places where life appeared on Earth. Warm ponds (primeval soups), mineral surfaces, underwater hydrothermal vents, volcanic lakes are only some among several proposals that have been put forth. As it is well known, in the middle of the last century Stanley Miller was able to synthesize aminoacids in a reactor simulating physical conditions which were believed to resemble those of the primitive earth. After this breakthrough, the interest of the scientific community focused on trying to identify different chemicals and chemical reaction sets which might be able to collectively self-reproduce, the main candidates being polypeptides and nucleic acids (mostly RNA). The importance of catalysis was stressed, and the candidate catalysts have been (and still are) polypeptides, RNAs, metal ions and others. Note also that the old proposal by Svante Arrhenius, that (at least some of) the important chemicals have come to the earth from space, has been supported by recent discoveries of many organic compounds in meteorites, asteroids and comets.

These researches are very important, and it would be wonderful to be able to precisely identify where and how life appeared on earth. However, one should be prepared to the probable failure of this effort, since the first living forms date back to more than 3.5 billion years ago, so the traces of those first progenitors might have been lost. Think for example of how difficult it is to reconstruct the migrations of the predecessors of *homo sapiens*, which were macroscopic organisms who lived a few millions, or a few hundred thousands years ago. Yet, the discovery of a new fragment of a bone in Asia or Europe may strongly affect the whole picture and its timing. If we think of small creatures, without bones that can become fossils, which lived three billion years ago, it is difficult to bet on the chances to follow their detailed pathways.

Moreover, even if we knew what the right molecules were, we might still be far from grasping what is life and how it came into existence. The point is that a set of chemicals is not yet a form of life, even if it is able to collectively self-replicate. You can take all the chemicals which are found in a bacterium, put them in a (very small) bottle—and nothing happens. The key point is that life requires a quite peculiar form of organization of its materials (mostly, organic molecules) and of their interactions. All the existing life forms are made of cells so, in order to understand how life emerged, we need to understand how cells appeared, and how they developed the ability to orchestrate the interactions which led to growth and to duplication with inheritance.

Surprisingly enough, there are some spontaneous physical phenomena which provide suggestions as to how this might have happened. It is well known that, under some suitable conditions, lipids in water spontaneously form vesicles, where an approximately spherical surface, composed by a lipid bilayer, surrounds a portion of the water phase. The structure of the bilayer resembles cell membranes, and it has sometimes been observed that these vesicles can undergo fission, a process that is at least superficially reminiscent of biological reproduction. The true picture is more complicated, and duplication of the vesicle does not guarantee duplication of its "genetic material". However in a series of papers (see Serra & Villani, 2017 for a comprehensive synthesis) we have been able to show that the two processes can spontaneously synchronize in a large number of cases (i.e. under different hypotheses concerning the type of chemical reactions and the architecture of the protocell). I will not discuss these quite technical aspects here but I will stress that synchronization is a condition which allows sustainable growth of a population of protocells, and that it allows Darwinian selection to take place. Note also that without membranes, without protocells, there are no individuals, no elementary units of selection for nature to operate on.

## 2 Protocells

Let us then take protocells seriously. While different and more sophisticated settings might be examined, we will consider here (one of) the simplest scenario(s) we can envisage, by simply supposing that

1. protocells are composed by lipid vesicles in an aqueous environment which contains some polymers and other chemicals

2. they form spontaneously by surrounding a portion of the fluid environment with a lipid bilayer; the chemical composition in the internal water phase is, on average, the same as that of the corresponding portion of the aqueous environment
3. the polymers in the water phases (either internal or external) can undergo several chemical reactions, using other polymers or simpler molecules available in the external environment (the “food”)
4. some reactions may allow the self-replication of a set of polymers
5. the protocell membranes are selectively semi-permeable, i.e. they allow the passage of some types molecules from the internal water phase to the external aqueous environment, and vice versa—while they cannot be crossed by other types of molecules
6. food molecules can freely cross the membranes (this hypothesis is not necessary, but it simplifies the description of the phenomena)

Let us first look at the external environment. The spontaneous formation of self-reproducing sets of molecules in the bulk of a reaction vessel is a highly unlikely event, which has been sometimes observed in experiments carefully engineered by smart chemists. But in the primitive earth there was some shortage of smart chemists, so we will consider the case where such spontaneous self-reproduction is not observed in external water phase.

On the other hand, the chemistry (i.e. the set of chemical species which are present, and the set of their reactions) must in principle allow for self-replication, otherwise we would never have life (and of course we know that life eventually emerged). There may be at least two main reasons why it is not observed in the large external environment: (1) the concentrations of the relevant chemicals may be too low, so that the relevant reactions are too slow, or (2) some species which are present inhibit some reactions which are necessary for self-replication to occur (e.g. by degrading some reactants).

It is often claimed, without further discussion, that cells (or other compartments) are necessary to avoid the diffusion of the reaction products in the external environment. However, this is not a convincing argument: since we have assumed that the internal and external aqueous environments are the same, it follows that the same reactions which take place inside the protocell do take place also in the bulk environment, and viceversa.

So, we may wander what is the actual role of protocells in this case. One important possibility is that membranes affect the rates of some reactions, playing a role similar to catalysts. They do not need to be true catalysts like enzymes, but they might create a local physico-chemical environment, where some reactions are favoured e.g. by the alignment of some reactants. Thanks to their symmetries, membranes should equally favour reactions both on the internal and external side, but the results would be different, since in the latter case diffusion would quickly dilute newborn species. The possibility that membranes play such an active role is important, however it is overlooked in most scenarios. We will develop our reasoning without assuming any help from of this kind.

**Table 1** Number of molecules in a protocell

| | 1 M | 1 mM | 0.1 mM | 1 $\mu$M | 1 nM |
|---|---|---|---|---|---|
| Typical (1 $\mu^3$) | $10^8$ | $10^5$ | $10^4$ | $10^2$ | 0.1 |
| Small ($10^{-3}$ $\mu^3$) | $10^5$ | $10^2$ | 10 | 0.1 | $10^{-4}$ |

Expected number N of molecules of a given species in a single protocell; two different volumes are considered ($10^{-18}$ and $10^{-21}$ $m^3$), and five different concentrations (1 M meaning 1 mole per L). Reprinted from (Serra & Villani, 2017), with permission

Other possibilities might be taken into account (like e.g. the idea that the rates of some, but not all, reactions are modified in the interior of a very small protocell) but there is no convincing evidence that this would favour collective self-replication.

There is however a familiar phenomenon which must necessarily take place. If protocells are formed by surrounding a portion of the aqueous medium, then different protocells may have different internal chemical compositions, due to random fluctuations from one point in space to another one. These fluctuations take place also in the bulk liquid, where however diffusion counteracts their effects, so they are very short-lived. However, semipermeable barriers prevent outside diffusion of some polymers, so the chemical mix inside a protocell may well differ from that of a neighboring one, and from that of the average external environment.

So, the fact that self-replication does not take place in the bulk of the external environment does not prevent it from taking place inside a protocell, where some molecules may be present at an unusually high concentration, or some may be absent.

In a sense, it's all a matter of size. The relative importance of fluctuations is a decreasing function of the number of molecules, i.e. (given macroscopically uniform density) of the protocell volume. Linear dimensions of lipid vesicles typically range from 100 nm (0.1 $\mu$m) to 10 $\mu$m. It is difficult to estimate the concentrations of primitive molecules, but we can look at a range of alternatives (see Table 1).

Remembering that the relative effect of fluctuations scales (in the simplest Gaussian estimates) as $1/\sqrt{N}$, it is easy to ascertain that there may be relevant differences in the case of small molecules and dilute concentrations of chemicals. Think for example that, in the case of small protocells and a 1 $\mu$M concentration, only one protocell out of ten will contain on average (a single molecule of) a given species.

It is therefore clear that small protocells allow a great number of different "experiments" to take place simultaneously. The protocells that host the most successful sets of molecules (i.e. those which self-reproduce and stimulate the replication of the hosting vesicles at the highest rate) will come to dominate the population. The stage for Darwinian evolution will be set.

This is a very simple reasoning, which looks convincing but which might be flawed by the fact that the set of chemical species and reactions may change in time. The complete reaction network comprises all the possible chemicals and all their reactions. Starting from a given initial condition, new species can be synthesized, starting from existing species and from the food, which can freely cross the cell membrane. The synthesis of new species may in turn lead to the disappearance of other species, therefore the internal composition of a protocell can change in time. If the processes were such that all the initial conditions would lead to the same composition (or to very similar compositions) the initial differences would be substantially irrelevant.

To tackle this issue, one needs to model the possible molecular types and their reactions. In order to do so, we have chosen a well-known case, the so-called binary polymer model (Farmer et al., 1986, see also Serra & Villani, 2017), where the molecules are linear strings (polymers) of binary symbols (monomers), and where the reactions are either condensations, where two strings are concatenated to form a larger one, or cleavages, where one string is cut into its two parts, choosing an arbitrary cut point. Only catalyzed reactions are assumed to take place at an appreciable rate, and polymers are chosen at random as catalysts of a given reaction. By extensive simulations it has been possible to show (Serra & Villani, 2019) that protocells which start from different initial compositions can lead to very different final compositions and, by coupling the concentration of some chemical species to the rate of reproduction of the protocells, to observe how some protocells can grow faster than others, thus allowing Darwinian dynamics.

While these results have been proven using a specific model of chemicals and reactions, it seems plausible that they can be generalized. This kind of studies indicates that the differential reproduction rates can be rooted in the random initial compositional differences among the primitive protocells (and of course in the nonlinear dynamics which amplifies such differences), thus hypothesizing that this type of randomness might be fundamental to the emergence of life.

## 3 A Shadow Biosphere?

The protocell models which we have considered belong to a broad class of mathematical or computational models which suggest that the emergence of some form of lifelike properties may be highly likely, under different assumptions concerning the chemical species which are involved and the physical environments where it is supposed to have taken place. It would be extremely satisfactory to be able to claim that life is an almost unavoidable outcome of this overall self-organizing activity, and to say, with Stuart Kauffman, that "we were expected", and we are not an accident in history (Kauffman, 1995). However, other authors, including (Monod, 1970) claim that life is an extremely unlikely event, so it is possible that it has happened only once in the universe. If this were the case, exobiology would be a useless waste of time.

However, we know that our earth can host life, as it actually does, so why should we not look for the existence of different life forms just here? If there is a high probability for life to emerge, it might have appeared several times on earth, and we should therefore be able to find the descendants of different predecessors. However, all known life forms are actually very similar to each other.

This may seem a surprising statement, since on a macroscopic scale we find "endless forms most beautiful", according to the famous quote by Charles Darwin (1859). However, when one looks at the microscopic level, the situation looks much more uniform. All living beings are composed of cells, where genetic information is stored in DNA (with the very peculiar exception of RNA viruses), proteins are synthesized in ribosomes using messenger RNA (obtained by transcription of DNA) and transfer RNA, and energy storage is mainly based on ATP. While L-forms and D-forms of optically active polymers are chemically equivalent, the chirality of nucleic acids and proteins is the same in all organisms. What is perhaps even more surprising, they also share the same genetic code: only 20 aminoacids are used in life, among the many more which exist, and the genetic code, i.e. the correspondence between bases in DNA or RNA and aminoacids, is always the same. The need for (at least) three bases per aminoacid can be inferred by the fact that the DNA code uses 4 different letters, so a code whose "words" used 2 bases might have coded at most 16 different aminoacids. But the correspondence of a specific triplet to an aminoacid seems quite arbitrary (although different hypotheses have been put forth), so it is startling that all organisms share the same code. The observation of this *a priori* unlikely homogeneity has led to the notion of LUCA (Last Universal Common Ancestor), i.e. a unique ancestor for all existing life forms.

But, if the emergence of life is "reasonably likely", and different types of life have been generated on earth, where have all the others gone? One frequent answer is that there has been worldwide competition, and the winner has destroyed the losers (either by "eating" them or by eating their food with higher efficiency). However, there are several niches on Earth, so it is conceivable that some different life forms might have survived somewhere, shielded from competition with "our" life.

Macroscopic organisms are probably unlikely to have passed unnoticed, but in the last decades very many single-cell organisms have been discovered, which can inhabit environments which were believed to be hostile to life, including those which can be found deep beneath the earth surface, or the ocean floor, or inside very hot or very acidic springs, or even inside higher animals and plants. According to some estimates, most biomass might reside in unicellular beings, which are very poorly known and understood (even the notion of species is not well-defined in bacteria).

These observations led to the hypothesis of the possible existence of a "shadow biosphere" on earth (see Davies, 2018), inhabited by these unconventional organisms, which are sometimes considered as "aliens", not because of their origin in outer space, but because of their belonging to a different genealogy. It is not entirely clear how to distinguish these aliens (say, a suspect microbial population) from still unknown forms of usual life, and there has been so far no proof of their existence, but their search opens interesting perspectives.

**Acknowledgments** Almost all my work on protocells has been designed and carried on with Marco Villani, and with important contributions from several colleagues, including Irene Poli, Norman Packard and Rudolf Fuchslin, and talented students, including Alessandro Filisetti, Chiara

Damiani and Alex Graudenzi. These works owe much to the support and inspiration of my friend Stuart Kauffman, and to the friendly and stimulating atmosphere that I found at the European Centre for Living Technology. Last but not least, everything started within the EU-funded project PACE (Programmable Artificial Cell Evolution).

## References

Darwin, C. (1859). *On the origin of life by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.

Davies, P. (2018). The demon in the machine. In *How hidden webs of information are solving the mystery of life*. Penguin Books.

Farmer, J. D., Kauffman, S. A., & Packard, N. H. (1986). Autocatalytic replication of polymers. *Physical Review D, 22*, 50–67.

Kauffman, S. A. (1995). *At home in the universe. The search for laws of self-organization and complexity*. Oxford University Press.

Monod, J. (1970). *Le hasard et la nécessité. Essai sur la philosophie naturelle de labiologie moderne*. Editions du Seuil.

Serra, R., & Villani, M. (2017). *Modelling protocells. The emergent synchronization of reproduction and molecular replication*. Springer.

Serra, R., & Villani, M. (2019). Sustainable growth and synchronization in protocell models. *Life, 9*, 68. https://doi.org/10.3390/life9030068

# Deep Learning and Knowledge Generalization

Guido Tascini

**Abstract** This work concerns the new studies related to the deep learning of machines. In particular, it tries to see what lies behind the behavior of deep neural networks, which have been very successful in various fields, such as image recognition, interpretation of natural language and much more. The work analyzes the deep networks and the behavior of the backpropagation algorithm. From this it seems that the success of these networks, which amazed the authors of the algorithms themselves, seems to lie in the processing of information and in the ability of the algorithms to extract relevant knowledge, discarding that which is not relevant for the purposes of the learning target. The bottleneck principle (Tishby & Zaslavsky, 2015 IEEE Information Theory Workshop (ITW), Jerusalem, pp. 1–5, 2015), in particular, appears to be a promising vision for the design of deep artificial neural networks, based on a general principle related to the processing of knowledge.

**Keywords** Autonomous knowledge learning · Backpropagation · Deep artificial neural network · Deep learning · Input data compression · Output prediction · Knowledge generalization · Relevant knowledge extraction · Shallow artificial neural network

## 1 Introduction

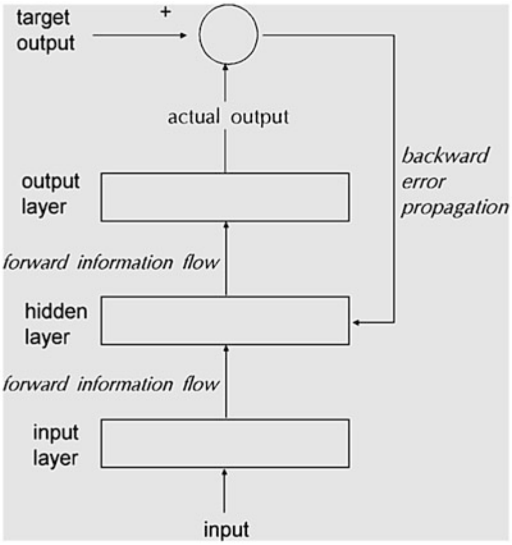
A Shallow Neural Network, has a single hidden layer, between an input layer and an output layer. The algorithm that associates the set of input patterns to the set of output patterns, named back-propagation algorithm, derives its name from the backward propagation of the errors on the output units: difference between real outputs and expected outputs. If the network has more than one hidden layers, we are talking about deep networks.

---

G. Tascini (✉)
ISAR-Lab, ITPI, Fermo, Italy

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_12

**Fig. 1** Backpropagation algorithm scheme



The scheme of the Backpropagation algorithm, for the shallow network, is represented in Fig. 1. As it can be seen, the normal information flows forward, while the errors flow backward and the errors derive from the comparison between the output that the network must have (target) and the calculated output from the network (actual). The algorithm tries to minimize these errors by varying the weights of the links and stops when the target output substantially coincides with the actual output (minimum of errors).

**Weights Initialization** Normally weights and thresholds are set, at the beginning, equal to small random numbers. The activation level of the 'entry unit' is set by the instance; as the activation level $O_j$ of the hidden unit or the output unit, it is determined by the expression:

$$O_j = F\left(\sum w_{ji} O_i - \theta_j\right)$$

where

$$F(a) = \frac{1}{1 + e^{-a}}$$

**Weight Training** We start from the output layer and work backward on the hidden layers by recursively updating the weights with the relation:

$$w_{ji}\,(t+1) = w_{ji}(t) + \Delta w_{ji}$$

where the variation of weights is given by the expression delta:

$$\Delta w_{ji} = \eta \delta_j O_i$$

with η speed parameter. A term, called 'moment', is added to this variation to speed up convergence:

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_j O_i + \alpha \cdot \lfloor w_{ji}(t) - w_{ji}(t-1) \rfloor$$

where $0 < \alpha < 1$.

**$\delta_j$**, gradient of the error, is calculated with the expression:

$$\delta_j = O_j \left(1 - O_j\right) \sum_k \delta_k w_{kj}$$

where $\delta_k$ is the gradient of the error corresponding to the unit k to which a connection from unit j points.

The iterations are repeated until convergence. Then *backpropagation algorithm* for shallow networks, with only one hidden layer, is the following:

```
Initialize the weights randomly
Do {
    Initialize the global error $E = 0$;
    For each $(X_k, t_k) \in TS\}$ {
        Calculate $y_k$ and the $E_k$ error;
        Calculate the $\delta_j$ on the output layer:
        Calculate the $\delta_i$ on the hidden layer:
        Update the network weights: $\Delta w = \eta \delta x$:
        Update the global error: $E = E + E_k$;}
} while ($E < \varepsilon$);
```

Training is carried out in the following three phases.

**Learning** In this phase the training set patterns set $(X_k, y_{dk})$, $k = 1, \ldots M\}$ and weights modified according to the Error-Backpropagation rule. It is important to choose the patterns of the training set well so that they are as representative as possible of the information that the network has to learn.

Such patterns can be presented:

*In a Bach* (or cumulative) mode. All the patterns are presented first, the error committed on each one is calculated, the error is added up and then the connection coefficients are modified;

*In a on-line* mode in which the connection coefficient values are updated after the presentation of each single pattern of the training set. Convergence occurs when it is reached a reduction of the global error, $E = \Sigma_k E_k$, so that the weights adapt to the input pattern: in practice so that it becomes $E < \varepsilon$.

**Generalization** A well-trained network must be able to generalize information. In the learning phase, after minimizing the errors committed at the exit, the weights are

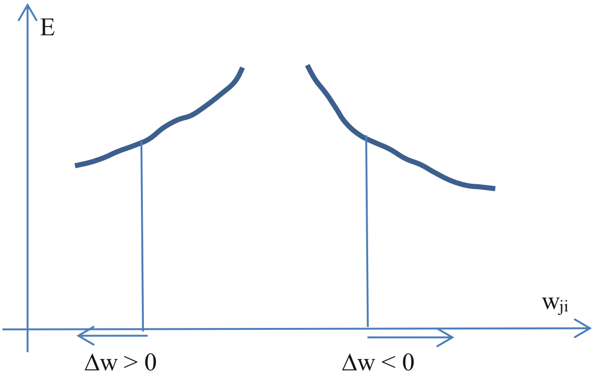


**Fig. 2** Gradient Descent method

frozen to proceed to the generalization phase in which the network responds well to examples never seen before.

**Convergence** The 'gradient descent' method is universally adopted for the convergence phase. Conceptually, the method consists in reducing the global error going down towards the minimum, along the curve E = f (w), with the calculation of the gradient.:

- if the gradient, $\partial E/\partial w_{ji}$ is positive, you must go towards the decrease of the weights (Δw <0),
- if the gradient, $\partial E/\partial w_{ji}$ is negative, you need to go towards weight gain (Δw > 0). See Fig. 2.

The error is calculated every time a training pattern is presented to the network and then a descent towards the minimum is performed along the curve E = f (w) following the decrease in the gradient. And there will be a gradient for each weight.

**Extended Delta Rule** y1 (actual output) is compared with y1 *(expected output) and the coefficients are increased by $\Delta w_{ij}$:

$\Delta w_{ij} = \eta \partial E/\partial w_{ij}$, with η = learning parameter. If as a measure of the error we have.

$$E = \frac{1}{2}\sum_i \left(y_i - y_i^*\right)^2$$

It leads to the explicit formula:

$$\Delta w_{ij} = -\eta \left(y_i - y_i^*\right) \frac{\partial F\left(P_i\right)}{\partial P_i} x_j$$

While if we are dealing with a finite set of training patterns it is more convenient to use the GLOBAL error:

$$E = \frac{1}{2} \sum_k \sum_i \left( y_{ik} - y^*_{ik} \right)^2$$

That give the *Extended delta rule.*

$$\Delta w_{ij} = -\eta \sum_k \left[ \left( y_{ik} - y^*_{ik} \right) \frac{\partial F\left(P_{ik}\right)}{\partial P_{ik}} x_{jk} \right]$$

In this last case, it is equivalent to the search for a local minimum of the value of E moving in the direction of the maximum decrease (gradient method).

## 2 Deep Neural Networks

Deep Learning (DL) is a branch of Machine Learning. It allows you to extract very complex information from a set of data, making it possible to carry out very complicated tasks, such as those related to the perceptual sphere. Deep learning models have the characteristic of being made up of different processing layers, each of which extracts a representation of the previous layer.

In the context of supervised deep learning, the most used class of models is the multi-layer neural network, or deep neural network (DNN). So it is a type of network built model, the main components of which are nodes, or neurons. As known, there are different classes of neural networks, depending on the type of nodes, and how they are connected to each other. The neural networks, on the basis of which the types of networks used in deep learning have been developed, are *feed-forward neural networks* (*FFNN*), whose operation is normally based on the "Back-Propagation" algorithm.

We can define the *FFNN* as follows: a network in which, if we number the vertices, all the connections go from one vertex to another of greater number. In practice the vertices are grouped into layers, and the connections go only from one layer to the higher layers.

The layers of the nodes form a hierarchical structure: the lowest layer is the *input* layer; the highest is the *output* layer. All the layers located inside are called *hidden* layers; see Fig. 3.

The Deep Neural Networks, with multiple layers of neurons, of the type feed-foreward, with many more then two hidden layers, and accelerated by the use of GPUs, have recently seen enormous successes in many fields. They have passed the previous state of the art in speech recognition, object recognition, images, linguistic modeling and translation.

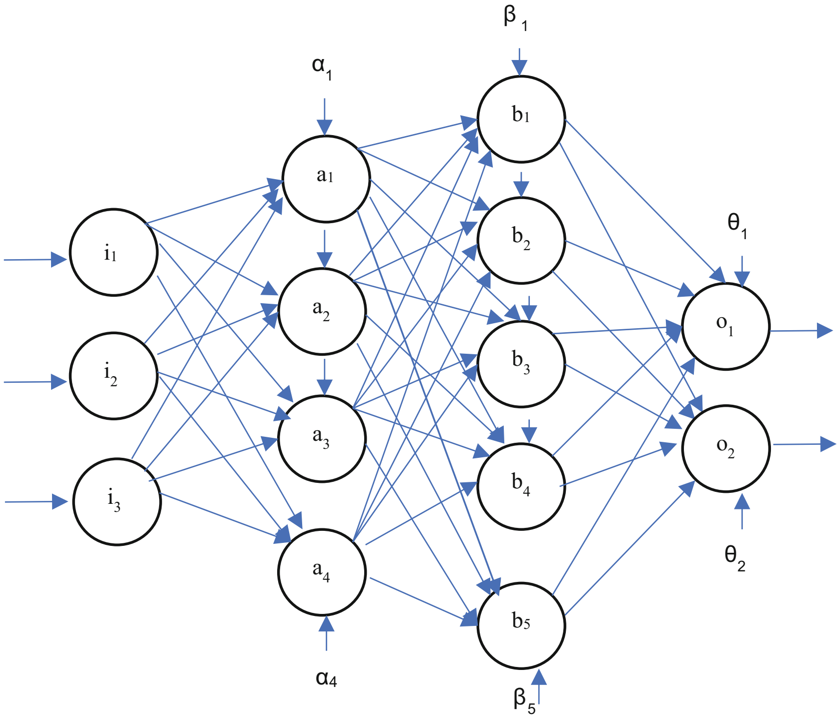


**Fig. 3** Deep Neural Network, of the type *feed-forward,* described by the sequence 3–4–5-2, with four layers: input layer, two hidden layers, output layer

The Fig. 3 illustrates a *deep neural network* with only two hidden layers. The show nn has three inputs (**$i_1$, $i_2$, $i_3$**), a first hidden layer ("**A**") with four neurons, a second hidden layer ("**B**") with five neurons and two outputs ($O_1$, $O_2$), that may be described by the sequence **3–4–5-2.** This network requires a total of (**3** * 4) weights +4 *bias* + (**4** * 5) weights +5 bias + (**5** * 2) weights +2 bias = **42** weights and 11 bias.

The example use as activation function the *hyperbolic tangent* for the outputs of the two hidden layers and the *softmax* for the output of the network. Then the formulas that calculate the feed-forward are as follows:

$A_i = \tanh(i_1p_{1i} + i_2p_{2i} + i_3p_{3i} + \alpha_i)$—first hidden layer,
$B_i = \tanh(A_1p_{1i} + A_2p_{2i} + A_3p_{3i} + \alpha_{i)}$—second hidden layer,
$O_i = \mathrm{softmax}(B_1p_{1i} + B_2p_{2i} + B_3p_{3i} + B_3p_{3i} + \beta_{\ i)}$—outputs.

The training standard of deep NN uses back-propagation algorithm. The deep neural network training, with multiple hidden layers, is more difficult than the *shallow neural network* training with a single layer of hidden nodes. This factor is the main obstacle to overcome in order to process networks with many hidden layers.

The connections, represented by arcs, are unidirectional and connect only nodes of one layer with those of the next layer. Each arc is associated with a parameter, called weight. In the initial modeling the arcs represented the synapses, that is, nerve impulses that are transmitted from one neuron to another and the purpose of these models was to identify which neurons were crossed by a sufficiently intense signal, omitting neurons, whose signal was below a certain threshold. We present the relationship between the layers of the network as a univariate relationship. For this we define:

- L: number of layers of the network, consisting of an input layer, an output layer and L—2 hidden layers;
- $p_1$: number of input nodes;
- $p_l$: number of nodes present in the l-th layer;
- $x_i$: value of the i-th input node;
- •$a_j^{(l)}$: value of the j-th node of the l-th layer;
- $w^{(l)}_{ij}$: coefficient associated with the arc that connects the i-th node of the l-th layer with the j-th node of the (l + 1) -th layer;
- $y_k$: value of the k-th output node.

The relationship between the input layer and the first hidden layer is:

$$z_j^{(2)} = w_{0j}^{(1)} + \sum_{i=1}^{p_1} w_{ij}^{(1)} x_i,$$

$$a_j^{(2)} = g^{(2)}\left(z_j^{(2)}\right).$$

Note how the j-th node of the first hidden layer takes on a value equal to $g^{(2)}(z^{(2)})$, where $g^{(2)}(\cdot)$ is a non-linear function, called activation function, while $z^{(2)}$ is the linear combination of the input nodes and the parameters $w^{(1)}$. To this linear combination is added the term:

$$w_{0j}^{(l)}$$

that is the parameter associated with the arc that connects a constant node equal to 1 with the j-th node of the (l + 1)-th layer.

This quantity acts as an intercept in the linear combination, and is introduced to model any distortion.

The relationship between the (l–1)-th layer and the l-th layer is defined as:

$$z_j^{(l)} = w_{0j}^{(l-1)} + \sum_{i=1}^{p_{l-1}} w_{ij}^{(l-1)} a_i^{(l-1)},$$

$$a_j^{(l)} = g^{(l)}\left(z_j^{(l)}\right). \tag{1}$$

The activation function $g^{(l)}(\cdot)$ is specific for the *l-th* layer, although a single activation function g $(\cdot)$ common in all layers is often used for the entire network. Finally, the output layer is produced through the relationship between the *(L-1) -th* layer and the following *L-th* layer:

$$z_k^{(L)} = w_{0k}^{(L-1)} + \sum_{i=1}^{p_{L-1}} w_{ik}^{(L-1)} a_i^{(L-1)},$$

$$y_k = g^{(L)}\left(z_k^{(L)}\right).$$

Both the number of output nodes K and the transformation function $g^{(L)}(\cdot)$ depend on the problem in question. For a unchanged regression problem, there is typically only one output node, therefore $K = 1$, while, a suitable choice of transformation function is the identity function, $g^{(L)}(z^{(L)}) = z^{(L)}$. For a classification problem, the number of nodes K coincides with the number of classes of the response variable that you want to model. Each node *k* indicates the probability of belonging to the *k-th* class. As a transformation function, it is often convenient to use the multinomial logistic function,

$$g^{(L)}\left(z_k^{(L)}\right) = \frac{e^{z_k^{(L)}}}{\sum_{j=1}^{K} e^{z_j^{(L)}}}$$

which is called the *softmax* function. Now ask:

$$\mathbf{a}^{(1)} = x = \left[1\ x_1 \ldots x_{p_1}\right]^T;$$

$$\mathbf{a}^{(l)} = \left[1\ a_1^{(l)} \ldots a_{p_l}^{(l)}\right]^T;$$

$$\mathbf{w}_j^{(l)} = \left[w_{0j}^{(l)}\ w_{1j}^{(l)} \ldots w_{p_l j}^{(l)}\right]^T;$$

$$W^{(l)} = \left[\mathbf{w}_0^{(l)}\ \mathbf{w}_1^{(l)} \ldots \mathbf{w}_{p_{l+1}}^{(l)}\right]^T;$$

$$\mathbf{W} = \left[W^{(1)}W^{(2)}\dots W^{(L)}\right];$$

$$\boldsymbol{y} = [y_1 \dots y_K]^T.$$

**Vector Notation** Adopting vector notation makes it easier and more intuitive formulate the relationship between two generic layers of the network:

$$\mathbf{z}^{(l)} = W^{(l-1)}\mathbf{a}^{(l-1)}, \tag{2}$$

$$\mathbf{a}^{(l)} = g^{(l)}\left(\mathbf{z}^{(l)}\right), \tag{3}$$

where the function $g\ (l)\ (\cdot)$ is applied element by element to the vector $z^{(l)}$. Consequently, the complete relationship between the input vector $x$ and the output vector $y$ is the following:

$$\mathbf{y} = \boldsymbol{f}\,(\mathbf{x};\mathbf{W}) = \boldsymbol{g}^{(L)}\left(\boldsymbol{W}^{(L-1)}\boldsymbol{g}^{(L-1)}\left(\cdots W^{(2)}\boldsymbol{g}^{(2)}\left(W^{(1)}\mathbf{x}\right)\right)\right) \tag{4}$$

## 2.1 Calculation of Parameters Via Backpropagation

For *regression* problems, we generally have a quantitative response variable $y = (y_1, ..., y_n) \in \mathrm{R}^n$, while for *classification* problems we use a qualitative response variable $y = (y_1, ..., y_n) \in T\ (y)^n = \{t_1, ..., t_K\}^n$, where T (y) is the set of modalities that can assume y. Consider a whole of data, consisting of n observations, for each of which are detected p explanatory variables, $x_i = (x_{i1}, ..., x_{ip}) \in \mathrm{R}^p$.

We want to adapt a neural network to the set of data, with the minimization of a given loss function $L[y, f\ (x;\ \mathbf{W})]$. This is achieved by looking for those values of the parameters $\hat{\mathbf{W}}$, such that

Loss function to be minimized is chosen from the following:

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\arg\min}\left\{\frac{1}{n}\sum_{i=1}^{n} L\,[y_i, f\ (x_i;\ \mathbf{W})]\right\}. \tag{5}$$

For **regression** problems

*Mean square error*

$$\mathrm{MSE}\,(\mathbf{W}) = \frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - f\ (x_i;\ \mathbf{W}))^2;$$

*Root of the MSE*

$$\text{rMSE}\,(\mathbf{W}) = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} (y_i - f\,(x_i;\mathbf{W}))^2};$$

*Mean absolute error*

$$\text{MAE}\,(\mathbf{W}) = \frac{1}{n}\sum\nolimits_{i=1}^{n} |y_i - f\,(x_i;\mathbf{W})|\,.$$

For **classification** problems,
*Misclassification rate*

$$\text{H}\,(\mathbf{W}) = -\sum\nolimits_{i=1}^{n}\sum\nolimits_{k=1}^{K} y_{ik}\log f_k\,(x_i;\mathbf{W})$$

where $y_{ik} = 1$ if $y_i = t_k$, 0 otherwise. Then minimizing cross-entropy corresponds to maximizing the log-likelihood (Hastie et al., 2009). The algorithm most widely used to estimate and calculate neural networks, is the backpropagation algorithm, adapted to Deep Neural Networks (see Rumelhart et al. 1986).

## 2.2 Backpropagation Algorithm for Deep Neural Networks

1. Calculate the value of the node $a^{(l)}$ for each layer l = 2, ..., L, using the current values of W,
2. For the output layer l = L, calculate

$$\delta^{(L)} = \frac{\partial L\left[y_i, \hat{f}\,(x_i;\mathbf{W})\right]}{\partial \hat{f}\,(x_i;\mathbf{W})} \circ \dot{g}^{(L)}\left(\mathbf{z}^{(L)}\right); \tag{6}$$

3. For the hidden layers (l = L—1, ...,2) obtain

$$\delta^{(l)} = \left(W^{(l)'}\delta^{(l+1)}\right) \circ \dot{g}^{(l)}\left(\mathbf{z}^{(l)}\right); \tag{7}$$

4. Having $\delta_2,\ldots,\delta_L$ it Is Possible to Derive the Partial Derivatives with

$$\frac{\partial L\left[y_i, \hat{f}\left(x_i;\mathbf{W}\right)\right]}{\partial W^{(l)}} = \delta^{(l+1)}\mathbf{a}^{(l)'}; \tag{8}$$

5. Update the W parameters using the gradient descent;
6. Start over with a new iteration from step 1, using the new values for the W parameters.

This algorithm solve Eq. (8), with a low computational cost. Normally most of the numerical optimization algorithms are iterative and require the calculation of the gradient of the loss function with respect to the parameters, of first and second order.

Keeping in mind that a multi-layered neural network has a very high number of parameters, the computational cost of calculating the second order gradient becomes excessive. If L are the layers, the network has L matrices of parameters W (l), each of which contains $p_l \times p_l + 1$ coefficients, where the number of nodes $p_l$ can reach a few thousand. For each iteration of the algorithm, the calculation of the first gradient requires a number of operations equal to the number of coefficients, while the operations required for the calculation of the second degree gradient grow quadratically as the number of parameters increases.

The advantage of the backpropagation algorithm is that, on the one hand, it does not require the second order gradient, and on the other, it calculates the first gradient only in the last layer, and then propagates it backwards in the other layers.

The algorithm alternates, for a given observation (xi, yi), with i = 1, ..., n, two steps iteratively: with the step *forward* you get f^ (xì; W) through (4), keeping W fixed, while with the step *backwards* you get the gradients and the parameters are updated. In machine learning, each iteration is called an *epoch*.

In the step forward, the value of the nodes a (l) for each layer l = 2, ..., L is calculated, using the current values of W (point 1 of algorithm backpropagation). Through formula (4), it is possible to obtain all the values of the nodes, a (l), and of the linear combinations, z (l), saving the intermediate quantities in progress. It is therefore necessary to initialize the parameters with randomly chosen values, close to 0.

Then the step backwards develops. This includes a *propagation phase* (points 2–4) and an *update phase* (point 5). The purpose of the propagation step is to compute all the partial derivatives $\partial L[y_i, \hat{f}(x_i; \mathbf{W})]$, with respect to the parameters. In practice, the quantities $\delta_L, ..., \delta_2$ are obtained, useful for calculating the partial derivatives, in an iterative way. The generic $\delta_l$ must be calculated as $\partial L[y_i, \hat{f}(x_i; \mathbf{W})]$ with respect to $z^{(l)}$. The $\delta_L$ of the output layer can be calculated with the "chain rule"; in substance δL is calculated as follows:

$$\begin{aligned}\delta^{(L)} &= \frac{\partial L\left[y_i, \hat{f}(x_i;\mathbf{W})\right]}{\partial \mathbf{z}^{(L)}} \\ &= \frac{\partial L\left[y_i, \hat{f}(x_i;\mathbf{W})\right]}{\partial \hat{f}(x_i;\mathbf{W})} \frac{\partial \hat{f}(x_i;\mathbf{W})}{\partial \mathbf{z}^{(L)}} \\ &= \frac{\partial L\left[y_i, \hat{f}(x_i;\mathbf{W})\right]}{\partial \hat{f}(x_i;\mathbf{W})} \circ \dot{g}^{(L)}\left(\mathbf{z}^{(L)}\right),\end{aligned}$$

where$\dot{g}^{(L)}$ indicates the first derivative of $g^{(L)}(\mathbf{z}^{(L)})$, and is easily obtained by deriving the expression (3) for l = L; the symbol ° indicates the Hadamard product (element by element product).

The δ $^{(l)}$ of the generic layer l is obtained as follows:

$$\begin{aligned}\delta^{(l)} &= \frac{\partial L\left[y_i, \hat{f}(x_i;\mathbf{W})\right]}{\partial \mathbf{z}^{(L)}} \\ &= \frac{\partial L\left[y, \hat{f}(x_i;\mathbf{W})\right]}{\partial \mathbf{z}^{(l+1)}} \frac{\partial \mathbf{z}^{(l+1)}}{\partial \mathbf{a}^{(l)}} \frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{z}^{(l)}} \\ &= \delta^{(l+1)} \frac{\partial \mathbf{z}^{(l+1)}}{\partial \mathbf{a}^{(l)}} \frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{z}^{(l)}} \\ &= \left(W^{(l)'}\delta^{(l+1)}\right) \circ \dot{g}^{(l)}\left(\mathbf{z}^{(l)}\right),\end{aligned}$$

where

$$\frac{\partial \mathbf{z}^{(l+1)}}{\partial \mathbf{a}^{(l)}} = W^{(l)'}$$

is the first order gradient of (2). This expression correspond to (7) of the backpropagation algorithm and is named *backpropagation equation.*

Having $\delta_2,\ldots,\delta_L$ it is possible to derive the partial derivatives with

$$\frac{\partial L\left[y_i, f(x_i;\mathbf{W})\right]}{\partial W^{(l)}} = \frac{\partial L\left[y_i, f(x_i;\mathbf{W})\right]}{\partial \mathbf{z}^{(l+1)}} \frac{\partial \mathbf{z}^{(l+1)}}{\partial W^{(l)}} = \delta^{(l+1)}\mathbf{a}^{(l)'},$$

In the updating phase, the parameter values are modified by means of the gradient descent, which uniquely uses the first-order partial derivatives, calculated in the propagation phase. The descent of the gradient is a numerical optimization technique that allows to find the minimum point of a function, using only the first derivatives.

Then the algorithm is restarted with a new iteration, using the new values for the W parameters.

## 3 The Gradient Descent

Let's now see the updating of the parameters, carried out through the descent of the gradient, which is what happens in point *5* of backpropagation algorithm. The gradient descent, based on the delta rule, is the most common and immediate method for updating the W (l) parameters (point *5* of algorithm) (Bengio, 2012). In this case, the updating of the parameters, at step t, takes place according to the Formula

$$W_{t+1}^{(l)} = W_t^{(l)} - \eta \cdot \Delta L\left(W_t^{(l)}; x, y\right), \quad \text{per} \quad l = 1, \ldots, L-1$$

where

$$\Delta L\left(W_t^{(l)}; x, y\right)$$

Is the gradient respect to $W_t$(i) of the argument of expression (5), that is gradient of

$$\frac{1}{n}\sum_{i=1}^{n} L = [y_i, f(x_i; W)]$$

the

$$\Delta L\left(W_t^{(l)}; x, y\right)$$

corresponds to

$$\Delta L\left(W_t^{(l)}; x, y\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial L[y_i, f(x_i; W)]}{\partial W_t^{(l)}}. \tag{9}$$

Essentially, if the gradient is negative, the loss function at that point is decreasing, which means that the parameter has to move towards larger values to reach a minimum point. Conversely, if the gradient is positive, the parameters have to shift towards smaller values to reach lower values of the loss function. The parameter $\eta \in (0, 1]$ is called the *learning rate*, and it determines the magnitude of the displacement.

### *3.1 Mini Batch Gradient Descent*

The previous method has several problems and limitations when applied to multilayered neural networks. The use of all data to perform a single update step involves considerable computational costs and greatly slows down the estimation procedure. Furthermore, it is not possible to estimate the model if the dataset is too large and cannot be loaded entirely into memory. In this regard, the mini-batch gradient descent technique is introduced. This consists in dividing the dataset into subsamples of fixed number mxn, after a random permutation of the entire data set.

The update is then implemented using each of these subsets, through the formula

$$W_{t+1}^{(l)} = W_t^{(l)} - \eta \cdot \Delta L\left(W_t^{(l)}; x^{(i:i+m)}, y^{(i:i+m)}\right),$$

where (*i: i* + *m*) is the index to refer the observation subset from *i-th* to *(i* + *m) th.* Then, for each epoch, instead of a single updating (with all data) they are done many updatings (mini-batch) by using the mini-batch data.

Advantages of this technique are:

- With little part of observations it is possible to meet better minima.
- The algorthm steps are so much faster and this fact guarantees a fastest convergence towards the minimum point.

**The learning rate problem:** setting too small values can lead to a very slow convergence, while large values can make the parameters fluctuate around the minimum without bringing the algorithm to convergence. Furthermore, dealing with this quantity with classic regularization methods (such as cross-validation) can be computationally too expensive. Finally, it seems inappropriate to think that all parameters need the same learning rate value to converge optimally. The problem of entrapment in local minima far from the absolute minimum. Since the models covered are highly parameterized, the loss functions previous discussed are generally convex in f (x; W), but not in W. This means that L [y, f (x; W)] has a single point of minimum for f (x; W), which is obviously the absolute minimum. Conversely, L [y, f (x; W)] has several local minima for W, of which only one is absolute. Then solve Eq. (5) and find the absolute minimum for W is somewhat complex, due to the high risk of obtaining a local minimum (Hastie et al., 2009). The attempt to solve the aforementioned problems allowed the development of subsequent improvements with the mini-batch gradient descent (Duchi et al., 2014).

$$w_{t+1,ij}^{(l)} = w_{t,ij}^{(l)} - \frac{\eta}{\sqrt{G_{t,ij} + \varepsilon}} \cdot g_{t,ij},$$

where $G_{t,ij}$ is the sum of the squares of the gradients with respect to $w_{t,\,ij}$ $^{(1)}$, up to time t, that is $G_{t,ij} = \Sigma_1{}^T (g_{t,ij})^2$. $\varepsilon$ instead is a is a smoothing term that serves to avoid a null term in denominator, and is usually set to values of order of $10^{-8}$. This allows to avoid the adjustment of the learning rate parameter, of which only an initial value is set, usually equal to 0.01.

Since, $G_{\mathrm{ij}}$ is a sum of positive terms, this quantity continues to increase with each epoch, and the learning rate decreases until it tends to 0. This problem can be solved by iteratively redefining $\mathrm{G}_{ij}$ as an average exponential mobile (EWMA). The mean at time t is then

$$\mathbf{E}\left[g^2\right]_{t,ij} = \gamma \mathbf{E}\left[g^2\right]_{t-1,ij} + (1-\gamma)\, g_{t,ij}^2,$$

where $\gamma$ is normally updated around 0.9.

The updating of the parameters therefore becomes

$$w_{t+1,ij}^{(l)} = w_{t,ij}^{(l)} - \frac{\eta}{\sqrt{\mathbf{E}\left[g^2\right]_{t,ij} + \varepsilon}} \cdot g_{t,ij}.$$

A further improvement is obtained by keeping in memory past values also of the term $g_{t,ij}$, and also applying an exponential moving average to the latter. This innovative method of gradient descent is called *Adam* (Kingma & Lei, 2015; Sebastian, 2016). It is determined with $m_{t,ij} = \mathbf{E}[g]_{t,ij}$ e $v_{t,ij} = \mathbf{E}[g^2]_{t,ij}$. The quantities are then defined

$$\begin{aligned} m_{t,ij} &= \beta_1 m_{t-1,ij} + (1-\beta_1)\, g_{t,ij}, \\ v_{t,ij} &= \beta_2 v_{t-1,ij} + (1-\beta_2)\, g_{t,ij}^2, \end{aligned}$$

where $m_{0,ij}$ *and* $v_{0,ij}$ are initialized to 0. and it is shown the correction:

$$\tilde{m}_{t,ij} = \frac{m_{t,ij}}{1-\beta_1^t}, \quad \tilde{v}_{t,ij} = \frac{v_{t,ij}}{1-\beta_2^t}.$$

Then the parameter updated becomes:

$$w_{t+,ij}^{(l)} = w_{t,ij}^{(l)} = \frac{\eta}{\sqrt{\tilde{m}_{t,ij} + \varepsilon}} \cdot \tilde{v}_{t,ij}, \tag{10}$$

with $\beta_{1,}$ $\beta_2$ that must have values, respectively, 0.9 and 0.999. The method appears very efficient.

## 4 Deep Neural Networks and Convolutional Neural Networks

Deep neural networks are more difficult to train than shallow neural networks. On the other hand, deep networks are much more powerful than flat networks (Goodfellow et al., 2016). A widely used type of deep network is the convolutional deep neural network (CDNN).

Starting from shallow networks, through many iterations, we can build ever more powerful networks. The techniques to be inserted later are: *convolutions*, *pooling* and *GPU* (LeCun et al., 2015; Ronen & Shamir, 2015). To this we add the *algorithmic expansion of data training* to reduce overfitting, the use of the *dropout technique* (Srivastava et al., 2014) and network composition. Let's consider as an example: Manuscript classification, using figures from the MNIST datase.

Starting with convolutional networks (Delalleau & Bengio, 2011) with shallow networks, through successive iterations, we gradually build more complex networks:

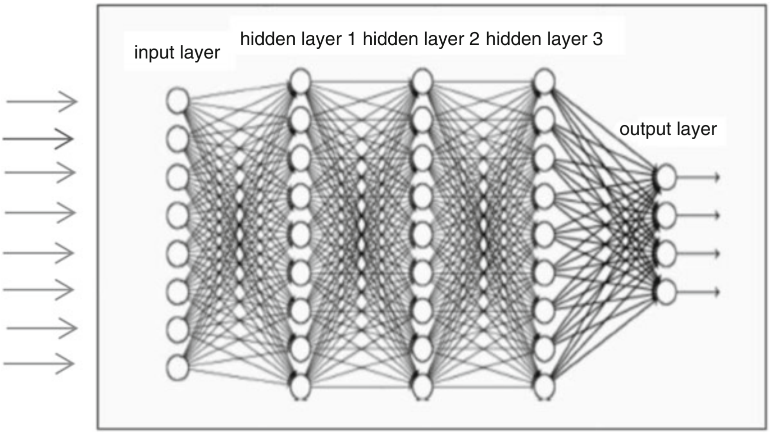


**Fig. 4** Convolutional N N, with: one input layer, three hidden layers, one output layer

The result will be a system that offers performance close to human. We will use the images not seen during training for the generalization test.

There have been spectacular recent advances in image recognition with convolutional networks; and also with recurrent neural networks, long- and short-term memory units, models that can be applied in speech recognition and natural language processing (Nielsen, 2015).

## *4.1 Convolutional Networks*

Here we have image recognition using networks with adjacent layers completely connected to each other (Krizhevsky et al., 2012). That is, every neuron in the network is connected to every neuron in the adjacent layers: Three basic ideas apply in convolutional neural networks: *local receptive fields, shared weights, and pools*. The input comes from squares of neurons, whose values correspond to the intensity of the pixels we are using (Fig. 4).

These squares are located in regions of the input image. Basically each neuron in the first hidden layer is connected to a small region of the input neurons, This region in the input image is called the *local receptive field*. Let's start with the top left corner and by scrolling the local receptive field over the entire input image we will have a different hidden neuron 'i' for each local receptive field (Fig. 5).

Steps greater than '1' and a direction different from the horizontal can be used. Shared weights and forecasts: each hidden neuron has a bias and weights connected to its local receptive field. We will use the same weights and biases for each of the hidden neurons. In practice, for the n.th hidden neuron, the output is: The use of the receptive field does not alter the recognizability of the image. The

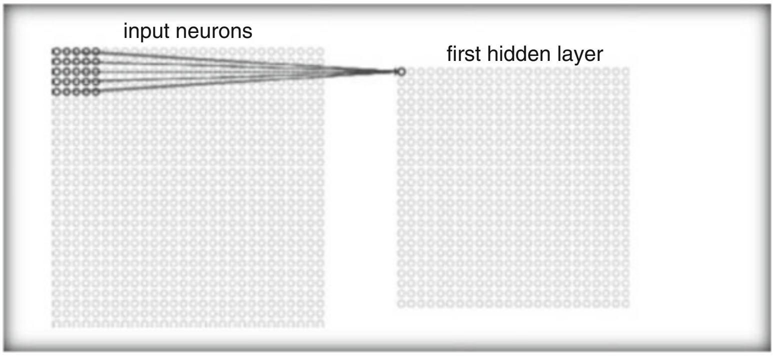


**Fig. 5** Receptive field connected to the first hidden layer

translation invariance of images also applies: the map from the input layer to the hidden layer is the feature map. We call the weights that define the characteristics in the map shared weights. The bias that defines the shared bias map file. The network map just described concerns only a localized feature (functionality). Image recognition requires multiple feature maps, so a full convolutional layer consists of several feature maps. Each map is defined by a set of shared weights and a single shared bias. The network can detect different types of feature-files, and each feature is detectable on the whole image. The images correspond to different feature maps (or filters). Each map is represented as a block image, corresponding to the weights in the local receptive field. Feature map example (see Fig. 7): The lighter blocks correspond to a smaller weight and the feature map responds less to the corresponding input pixels. Darker corresponds to greater weight, and the feature map responds more to corresponding input pixels.

Intuitively, it seems likely that the use of the translation invariance by the convolutional layer will reduce the number of parameters required to obtain the same performance as a fully connected model. This will also result in a faster workout. Intuitively, it seems likely that the use of the translation invariance by the convolutional layer reduces the number of parameters required to obtain the same performance as a fully connected model. This will also result in a faster training (Fig. 6).

**Pooling Layers** Pooling layers are placed immediately after the convolutional layers. The pooling layers simplify the information file that exits the convolutional layer: a pooling layer takes the output of each map of the characteristics of the convolutional layer and creates another map of condensed features.

For example, it condenses a region in the previous layer. Common procedure for pooling is *max-pooling*: the pooling unit takes only the maximum activation value in the input region (Fig. 7).

Example: max-pooling applied to each of three feature maps (see Fig. 8). The convolutional and max-pooling layers are similar to Neural Networks for Deep Learning.

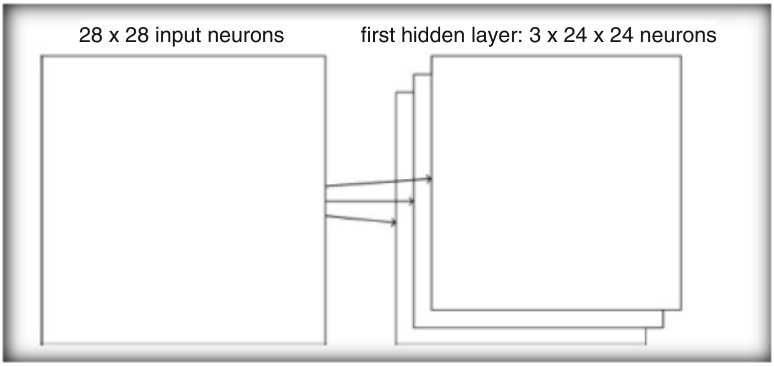


**Fig. 6** Input layer connected to three feature maps

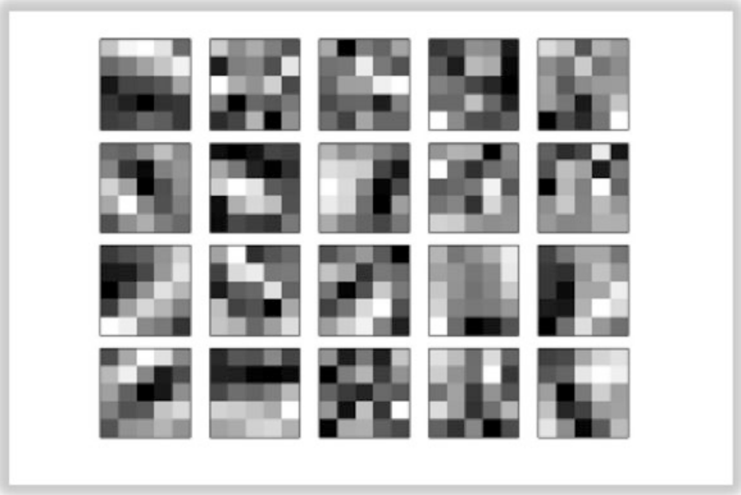


**Fig. 7** Feature map: block image, corresponding to the weights in the local receptive fields

**Using Rectified Linear Units** There are many ways to vary the network in an attempt to improve results.

For instance we can change neurons: ***instead of using the sigmoid activation, we use rectified linear units***. In practice We'll train for epochs. I also found soma advamtage by uinge some regularization, *with regularization parameter.*

**Expanding the Training Data** Another way to improve the results is by *algorithmically expanding the training data.* A simple way of expanding the training data is to ***displace each training image by a single pixel***, either up one pixel, down one pixel, left one pixel, or right one pixel. Using the expanded training data we can obtain a *better percent training accuracy*.

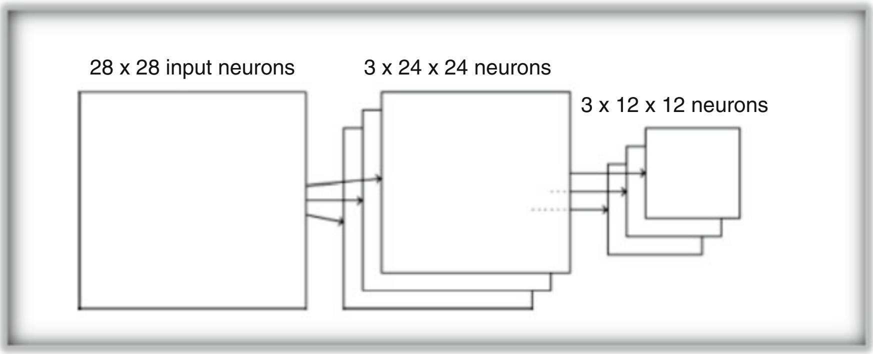


**Fig. 8** From Imput layer to 3 feature maps and then to 3 pooling maps

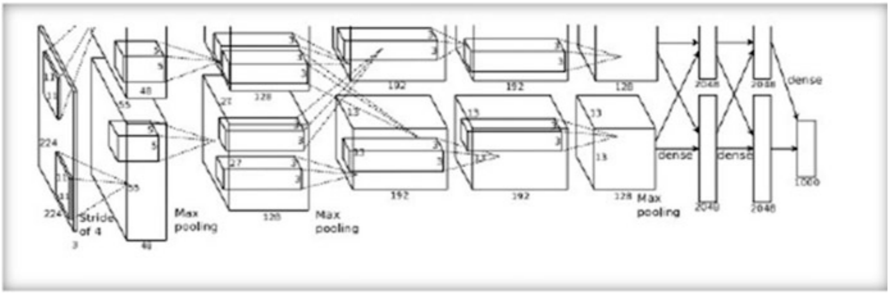


**Fig. 9** DCNN of Krizhevsky, Sutskever and Hinton

**Progress in Image Recognition** A best paper of Krizhevsky, Sutskever and Hinton appears in 2012 (Krizhevsky et al., 2012). **They** trained and tested a **DCNN** by a restricted subset of the ImageNet data. They used the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC-2012). The used competition dataset gave them the possibility of comparing their approach with others. The ILSVRC-2012 training set contained about 1.2 million ImageNet images, from 1000 categories. From the same 1000 categories performed validation and test sets containing, respectively, 50,000 and 150,000 images.

As an example of good architecture it is intresting to see the DCNN of Krizhevsky, Sutskever and Hinton.

The DCNN of Krizhevsky, Sutskever and Hinton has layers of hidden neurons.

*The first hidden layers* are *convolutional* layers and some with max-pooling, the *next layers* are *fully-connected* layers.

Note the layers split into 2 parts, corresponding to the 2 GPUs.

*The input layer* contains neurons, representing the RGB values for a image. ImageNet contains images of varying resolution, while a neural network's input layer is usually of a fixed size. The net dealt with this by rescaling each image so the shorter side had length .

The *first hidden* layer is a *convolutional layer*, with a max-pooling step. It uses 11 × 11 local receptive fields, and a stride length of 4 pixels. There are 96 feature maps, split into 48 feature maps on each *GPU.*

A *max-pooling* is in this and later layers, and done in 3 × 3 regions; pooling regions may be overlapped.

The *second hidden* layer is also *convolutional*, with a max pooling step. It uses 5 × 5 local receptive fields. There are 256 feature maps, split into 128 on each GPU.

The *input channels* are used *only by the feature maps*. This is because any single feature map only uses inputs from the same GPU.

The *third, fourth and fifth hidden* layers are also *convolutional*, but they do not involve max-pooling; their parameters are respectively:

- (3) 384 feature maps, with 3 × 3 local receptive fields, and 256 input channels;
- (4) 384 feature maps, with 3 × 3 local receptive fields, and 192 input channels;
- (5) 256 feature maps, with 3 × 3 local receptive fields, and 192 input channels.

The third layer involves some inter-GPU communication (see figure) so the feature maps use all 256 input channels.

The *sixth and seventh hidden* layers are *fully-connected* layers, with 4096 neurons in each layer.

***The output layer*** is a 1000-unit softmax layer.

## *4.2 Deep Learning and Knowledge Relevance*

The new era of Artificial Intelligence, linked to deep learning, was born with the overcoming of Expert Systems and the difficulties encountered in defining all the rules necessary to create a useful and efficient Expert System. In practice, the A.I. has gone from trying to provide the machine with the necessary knowledge, to making the machine learn this knowledge automatically. And this is how Machine Learning was born, and Deep Learning in its field, with the successes we know in the field of image recognition, speech, natural language, and in many other sectors in which Machine Learning is applicable.

In practice, the turning point took place by abandoning the design of systems that contained all the necessary knowledge for the intelligent machine, turning to the design of systems that independently learned the necessary knowledge. Machine Learning, after a period of interesting but not optimal results, has recently accelerated, thanks to progress in computer technology on the one hand, and to the development of decidedly efficient algorithms, based on innovative artificial neural networks, and, in this context, of Deep Learning. The singular aspect of this breakthrough is linked to the successes of these algorithms, whose dynamics and founding principles possessed dark sides and all to be investigated.

However, some glimmer is making its way. In particular by analyzing one of the best known and most effective algorithms: the Backpropagation Algorithm (Rumelhart et al., 1986; Shamir et al., 2010). The machine that learns to recognize things never seen before selects the information it treats based on its importance. The degree of importance of the information corresponds to its generalization. In practice, the machine that learns to recognize objects does so by evaluating the importance of the information that the object carries with it. In this regard, in the behavior of the Backpropagation Algorithm, we have seen just what has just been said. The algorithm in its iterations ends up filtering the unimportant information, and preserving the broader one, that is, of a general type. And therefore, after the training phase with the training set, the machine will be able to recognize objects never seen before. That is, the machine is able to generalize its knowledge.

The phases observed by Tishby and Zaslavsky (2015) during the run of backpropagation algorithm, in a deep network, can be summarized as follows:

*Initial state*: Layer 1 neurons encode everything about the input data including all information on its labels. In the higher layers, in which neurons are located, they are in almost random state, with little or no relationship to the data or their labels.

*Adaptation phase*. As the DL begins, neurons in the upper layers gain information on the input and get the best of adapting labels to it.

*Phase of change*. The layers suddenly change their behavior and begin to forget information about the input.

*Compression Phase*: the higher layers compress their representation of the input data, taking what is relevant to the output label. They take the best to predict the label.

*Balance between security and compression*. The last layer achieves a good balance, retaining only what is necessary to predict the label.

Naftali Tishby and others have analyzed deep neural networks and defined the ‘Information Bottleneck Principle’ (Tishby et al., 1999; Tishby & Zaslavsky, 2015).

In practice, this principle allow to reach the theoretical limits of the optimal information in the DNNs: that is, they say, obtain the generalization limits of finite samples. This is quantifiable both by the constrained generalization and by the simplicity of the network.

We can analise the compromise between the compression of input data (due to bottlneck) and the output layer that preserves the prediction of supervised target. Closely connected to this could be the optimal architecture of nn: layers number, characteristics, connections.

In their experiments, Tishby and Shwartz-Ziv monitored the amount of information each layer of a deep neural network held on the input data and the amount of information each held on the output label. The networks appear to converge at the theoretical limit of the information bottleneck: theoretical limit that represents the optimal system for extracting relevant information: the network appears to compress the input as much as possible without sacrificing the ability to accurately predict its label.

We can argue that this trade-off between input compression and output prediction can correspond to reducing (compressing) knowledge of the input, distinguishing what is not necessary, and is lost, and preserving what is relevant (general) for the output.

If this can be seen as more then the behaviour of some algorithms, but will become a general computational method, we would revolutionize the design of deep learning systems by designing their optimal architecture.

## References

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade* (pp. 437–478). Springer.

Delalleau, O., & Bengio, Y. (2011). Shallow vs. deep sum-product networks. In *Advances in neural information processing systems* (pp. 666–674).

Duchi, J. C., Jordan, M. I., Winwright, J., & Wibisono, A. (2014). Optimal rates for zero-order convex optimization:the power of two function evaluations. *arXiv:1312.2139v2.Mat oc.*

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Kingma, D. P., & Lei, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of ICLR 2015*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)* (pp. 1106–1114).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Nielsen, M. A. (2015). *Neural networks and deep learning*. Determination Press.

Ronen E., Shamir O. (2015). The power of depth for feedforward neural networks. *arXiv preprint*.

Rumelhart, D., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Sebastian, R. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint ar Xiv:*1609*.04747.*

Shamir, O., Sabato, S., & Thishby, N. (2010). Learning and generalization with the information bottelneck. *Theoretical Computer Science, 411*(29–30), 2696–2711.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958.

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. In *Proceeding of the 37th Annual Allerton Conference on Communication, Control and Computing* (pp. 368–377).

Tishby, N., & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW), Jerusalem* (pp. 1–5). https://doi.org/10.1109/ITW.2015.7133169

# Reasoning About Reason: Why Philosophy Should Now Abandon Monism in Favor of Pluralism

**Lucia Urbani Ulivi and Primavera Fisogni**

**Abstract** The term *reason* is used in a widespread and recurring way to indicate one of the traits that characterize human beings in a non-negotiable way. While this or that human behavior can be considered “irrational”, no human individual can be qualified as “irrational” *per se*, whereas a person can be “immoral” or “amoral”, devoid of aesthetic sense or incapable of linguistic expression, etc. Unlike other aspects, reason is always coextensive with humanity: as long as there is one, there is also the other. And this consideration already raises questions; it opens to reflections and calls for clarifications and explanations.

A classic starting move to dig into the theme is to look at the past, to the history of the term “reason”, and to proceed with a philological recognition. Such an approach is often useful in recovering not only the semantic outline, but also the conceptual groove through which the current meaning of a term has stabilized. This shift into the investigation of reason is a good point of departure, but honestly it does not suffice by itself. “Ratio”, the etymological antecedent of “reason”, provides chameleonic mutations over time, it intertwines and overlaps with “intellect”, “logic”, from which in different eras it sometimes diverges or converges, without a linear genealogical transmission. This historical complexity has still a lot to teach us, and it must be kept in mind, but it is also necessary to exploit contemporary cognitive resources to elaborate a concept of reason that is current and suitable for today’s world.

The aim of this essay is to make a contribution to philosophical research by making a proposal that does not operate by simplifying the problem and reducing it to a few components, but rather intercepts its richness and complexity, in order to achieve a contemporary concept of reason for our world, philosophical as well as cultural, making available to scientists and philosophers, sophisticated intellectuals

---

L. U. Ulivi (✉)
Former at Department of Philosophy, Catholic University of the Sacred Heart, Milan, Italy

FTL, Lugano, Switzerland
e-mail: lucia.ulivi@unicatt.it

P. Fisogni
Journalist, Responsible Editor Newspaper ‘La Provincia’, Cernobbio, Italy

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_13

and men of good will, a reference and a debating platform within which to interact with critical attention and intellectual honesty.

**Keywords** Abduction · Logic · Monism · Pluralism · Reason · System

## 1 The Meaning of "Reason"

### *1.1 A Brief: Though Almost Impossible—History*

A tangled term of the philosophical repertoire, the word "reason" presents chameleonic mutations over time. It intertwines and overlaps its meaning with "intellect" and "logic", from which in different times it sometimes diverges or converges, without maintaining a prevailing or stable semantic line.

Given all this, any effort of tracing the philological etymology of "reason" (which probably derives from the Latin *ratus*) is expected to be almost zero interest for philosophy. By trying to advance in such a complex subject matter, it may be promising the search for its conceptual antecedents, which can be found in a cluster of related terms provided by the classical Greek thought: *logos*, *noesis*, *dianoia*. Precisely *logos* displays the semantic density that characterizes every ancient utterance. In *logos* (from *legein*, to connect, to link) we face the very idea of a profound connection that binds things together, which can be grasped by human beings according to an immediate mental act, the *noesis*, which organizes the world scenario without any mediators. It becomes *dianoia* when the *noesis* is expressed in discursive or propositional form and eventually regimented in logical arguments.

If the story of "reason" is impossible to sketch in a very linear way, being destined to disperse itself in several semantic streams, the one that concerns its conceptual antecedents opens a broader perspective: it includes not only the discursive knowledge (*dianoia*), but also a wide number of references (*noesis*) starting from which the discourse is structured. The concept of *logos* also contains—at least in some classical authors among which Aristotle stands out—also an implicit but undeniable ontological commitment: human understanding, the *subjective logos*, can effectively perform its capacity to understand the world because the world consists of a rational structure, an *objective logos*, which makes it understandable.

If *logos* is the most plausible conceptual antecedent of "reason", it possesses, with respect to "reason", a semantic density that the latter no longer has. As if, we could otherwise say, in the historical path that led from *logos* to "reason" the concept has undergone a process that emptied it out by reducing it to a few traits. Do we have reasons that allow us to check this hypothesis? What has ever happened in the historical-semantic evolution that led from *logos* to *ratio*? What idea do we have today of reason, in philosophy and in common sense?

## 1.2 The Current Use of "Reason" in Philosophy and in Common Sense

Even if we are looking for the prevailing current use of the term "reason", it is necessary to immediately exclude that any current use exists absolutely, or that it can be traced without geographical, cultural, historical, personal, and many other limitations.

In contemporary European culture, among populations that share wide cultural spectra and a sufficiently homogeneous cultural past in which a semantics in good approximation uniform has settled, the utterance "current use" corresponds to the ordinary meaning provided by a good dictionary.

In this context, and within these semantic limitations, reason is primarily understood in term of reasoning, as the ability to correctly argue or "any process of drawing a conclusion from a set of premises" Blackburn 1994 (Oxford Dictionary of Philosophy, 1994: 320). Despite its precision, this definition does not allow to distinguish reason from the realm of logic, which is notoriously the discipline that specifies the conditions of correct inferences. The prevalent philosophical use follows the meaning ascribed by common sense; a rather extensive literature expresses with sophisticated specialist discourses (see, among many, Putnam, 1981; Simon, 1983; Rescher, 1988; Stich, 1991; Nozick, 1993) the basic idea that there are some universal standards of rationality. Such reduction of rationality to logic is sometimes tempered by modest concessions to inevitable but tolerable pragmatic deviations from the logical standards.

In short, not only for common sense, but also for philosophers, reason is mainly logical, and the deviations from formal correctness that are frequently found in the vast sea of human reasoning should be attributed to human irrationality, or to the particular circumstances that limit and distort the correct application of inference rules. (Piattelli Palmarini, 1994).

If we describe "reason" in terms of "logic", limiting the concept of reason within the boundaries of logic, we offer an easy conceptualization, simple to manage and apparently not problematic, of the term "reason". But have we answered the question posed before ("What the reason is?"), or do we have just ignored in what reason exceeds logic or is not superimposable to it? It is well known that simplifying ideas tend to successfully substitute more complex ones, but to better understand how such depletion took place, it is worth asking when this simplifying scheme was founded and established, given that not even Aristotle, the inventor of logic, would have probably subscribed to it. The turning point came in the seventeenth century, due to the New Mathematized Science that reinforced the objectivity and the certainty of its observations by expressing them in mathematical, formal or formalizable language.

Descartes (1641) took on the task of transforming the scientific method into a metaphysical scenario by codifying and transmitting many forms of philosophical reductionism such as mechanism, materialism, reductionism, mind-body separation to future generations. The Cartesian intellectual legacy imposes strict constraints on the subsequent research and orients its development through some powerful suggestions: the reason is univocal, it is nothing but formally correct reasoning, all the sciences use the same "rational" method, the errors of reasoning depend on the interference of the "irrational", generated in the heterogeneous spheres of will, passions, obstinate and rebellious unjustified beliefs. The identification of reason with logic became a paradigm, so pervasive to be implicitly accepted also by its opponents, who could escape the motto "Reason is nothing more than logic" only devoting themselves to "irrational" theories.

The conceptual constraints imposed by Descartes are, *de facto*, accepted and considered non-reviewable in many areas of philosophical thought up to the contemporary age (see the Philosophy of mind, and especially the debate on A.I., Scientism, Physicalism). The shared conceptual inheritance provided us with the idea that reason consists of correct argumentation and that the term "reason" is univocally referential. Furthermore, it is given for granted that philosophy and the sciences are qualified for common use of reason; any deviation from this use is to be considered as a sort of weakening. As a consequence of it, history, psychology, medicine, and all the arts that cannot satisfy the universal criteria valid for reason are rejected in those uncertain and cognitively opaque domains that we can only qualify as irrational.

Common sense and philosophy converge in identifying reason with logic despite the impossibility of proving the co-extension of the two domains and despite the innumerable corrections that are necessary to exclude logical incorrectness from the realm of reason as well as to exclude that logically correct arguments must be included in the domain of the non-rational (Devlin, 1997).

## *1.3 Reason and Logic*

Contemporary Anglophone thought has questioned the epistemological status of reason by asking whether rational standards of argumentation exist (Stich, 1991: 49).

A positive answer has two implications: first, the standards of rationality are specified in detail. Secondarily, reason is co-extended with logic because, if such standards exist, they correspond to universal criteria of logical correctness.

Putnam (1981: 104) wonders whether there is an ideal theory of rationality, which establishes the necessary and sufficient conditions for a belief to be considered rational in current circumstances and in all possible worlds. The problem is whether a criterial theory of rationality might exist. The answer can only be negative; Putnam shows that looking for such a criterial theory is a consequence of implicit scientism and reductionism but also, one might suggest, of a logicist assumption.

The search for universal standards of rationality has not provided valuable results and has forced many authors towards a pragmatism which by renouncing that claim, obtains in exchange the legitimacy of behaviors, decisions and choices, which are dominant and successful in our life, even though they cannot satisfy the requirements of any logical correctness. Pragmatism recognizes the appropriateness of the *pragma*, but it eludes the underlying problem because it does not address the twofold question: "in what relationship are logic and reason? Is it possible that reason can deviate from logic while maintaining the truth value?"

Piattelli Palmarini (1994) proves the logical inaccuracy of everyday reasoning and concludes for the unreliability of ordinary reason, which does not stand up to the logical test. He too seems to regret the formal inappropriateness of reason, without further pursuing the investigation.

Damasio (1994, 2010) and Devlin (1997) are the scientists who subjected the problem of the relationship between logic and reason to a drastic revision. Precisely they demolished the rationalist equivalence that, from Plato onwards, through Descartes to Turing (1992), Minsky (1986) and Winograd and Flores (1986) was unable to perceive the profound difference between describing human behavior in terms of rules or mathematical formalizations—a perfectly legitimate description in terms of the science that carries it out—and reducing the human behavior to such rules, neglecting the fact that actions are performed according to abilities, skills, moral constraints, preferences that escape formalization.

Thanks to Damasio and Devlin, philosophers have now the tools to get out of the logicist dilemma: either the reason is logical, or we are consigned to skepticism.

Damasio proves the embodiment of reason in opposition to the disembodied formality of logic and describes perfectly rational though illogical behaviors and completely irrational, but perfectly logical, behaviors. In the same line, Devlin criticizes logicism, that is to say the extension of logic to a univocal and unitary reference for human behavior and reasoning, and emphasizes how the context influences the standards of rationality—which undergo notable changes according to circumstances—and demonstrates the irreducibility of human actions to a repertoire of formal rules. Devlin also points out how much the meaning of the same term is influenced by the circumstances in which it is used and how much communication depends on the structure of conversation and culture.

In the light of Damasio's and Devlin's discoveries—just to name the most influential ones—it is necessary to reverse the logicist relationship between logic and reason: reason is the vast field that dominates human activity, of which logic is a subspecies; in some cases we use logic because we deem it appropriate and useful in a given context, in other cases we neglect or violate it openly without the rationality of what we do, say or think being weakened. If at the entrance to Macy's store we read: "on the escalator it is mandatory to carry dogs in your arms", an extremist logicist will go in search of a dog to carry in his arms, sacrificing context, circumstances and reason, where a human being who intends to keep in line with reason, even at the cost of sacrificing logic, will easily be riding on an escalator without a dog in his arms (the example is from Devlin, 1997: 270). As Devlin clearly remarks in his passionate investigation at the edges of logical thinking, it is not the logic and semantics of the sentences alone that can make their meaning fairly understood. In all the discourses the meaning is determined also by reference to the context given.

Here is an example: “A bachelor is an unmarried man”; this guy is a bachelor, so this guy is unmarried. An unmarried man can get married. But this guy can get married because although he already has three wives, he can get a fourth because he is Muslim, so it is not true that this guy is both bachelor and unmarried. Furthermore, this guy is a bachelor because he can get married, but he is not a bachelor because he is married. This guy is both a bachelor and a non-bachelor. In this case the problem is generated by the very meaning of “bachelor” which is understood as having a univocally referential meaning, that is, as a logical-linguistic symbol, while it undergoes important changes that depend on the context and circumstances of the utterance.

If the Platonic line has historically been dominant, the time has come to re-evaluate Aristotle’s philosophical attitude, according to whom logic is one of the ways of expressing the *logos*, but it is not the only one nor the best.

It is beyond the scope of this investigation to discuss the historical debate about reason and logic, we have simply limited to sketch some lines in order to introduce our theoretical claim.

## 2 A Philosophical Proposal

The etymological and semantic history of the term “reason” can supply useful insights, however, as we have seen before, it does not convey a concept capable of adequately supporting the many facets and features that belong to reason. It is necessary to venture into a new path and identify a definition that also takes into account the new horizons of knowledge opened by the sciences. Here is a first proposal, a deliberately broad one, therefore marked by a necessary level of vagueness and openness.

> Reason is the ontological principle underlying the structuring of reality and at the same time the epistemological counterpart that guides us in its understanding. It is the interface between the subject and the world.

We can try to better specify the meaning of reason and to deepen its understanding. Taken with reference to the human subject, reason is the ability to implement strategies for understanding the world and ourselves; it guides behaviors, it uses arguments, and it knows how to orient itself amid different hypotheses and scenarios. Rooted in the body and in contact with the emotional experience, with which it intertwines a silent and continuous confrontation, reason gives rise to different scenarios, evaluating the consequences and orienting towards choices and decisions.

Its action goes beyond what is present to conscious awareness and also includes subliminal levels in which it shows its activity by generating effective and efficient behaviors which remain unknown or ignorable to consciousness. Thanks to the bond that connects it in a profound and continuous way with the body, reason is embodied in a very full sense, as it moves within bodily constraints and limits. The body anchors it to the world of things, making it unrealistic to believe that the subject can disregard of the object and vice versa.

Epistemology and ontology cannot do without each other. Reason is the interface between the human subject and the world, it connects and puts them in communication for the aspects in which they diverge, and at the same time it makes them partakers of converging and common traits. Human beings can exercise their understanding of the world because both human beings and the world have a structure of reason which structures their configuration.

We therefore consider reason as the ontological principle that structures and organizes reality and as the epistemological counterpart that guides us in understanding ourselves and the world.

Reason is principle (*arché*), in the Aristotelian sense: "Ontological foundation of entities and gnoseological foundation of knowledge". Translated into a philosophical language for the contemporary context, reason is irreducible to another concept, in the sense that the notion of reason cannot be expressed in terms of other conceptualizations. It follows that one could speak of reason as a "primitive", a term dear to logicians, or even, in a classical way, as a "principle", in the sense of "what structures the constitution of an entity".

If we apply to reason the meaning of "principle" as seen above, we achieve a more articulated and less generic understanding of the trait that so typically identifies the human: reason constitutes an internal criterion that structures the ontology of the human world, gives shape and character to all its values and manifestations, and it is expressed in its phenomena. Used as a knowledge tool, it allows and supports cognitive activity. If we wish now to move from the maximum degree of the conceptual extension of "reason"—which inevitably involves the minimum level of intension—to its specifications, we will trace different relevant aspects according to the fields and interests of investigation. In this way, reason's specifications will gain in terms of intension what they lose in terms of extension.

## 3 One Reason, Many Viewpoints

### *3.1 From an Ontological Point of View Reason Can Be Subjective and Objective*

From an ontological point of view, there is an objective reason that structures reality, organizes it into distinct phenomena and objects, activates a processual activity that consists of the arising and the decaying of phenomena. The world, by being made up of separate entities—even if not isolated ones—is imbued with reason. From the ontological point of view, the peculiar structure of reason is inherent to the world, and would persist even in the absence of rational observers. This is not an epistemological hazard, but a condition rooted into ontological processuality.

In the world we are acquainted with, symmetries and symmetry breaks are dominating. In this scenario, phenomena endowed with properties and characteristics emerge. They do not only hold a certain degree of stability, but also changes according to dynamics specific to each domain.

With human beings, reason also acquires a subjective value: it becomes the tool that grasps and understands the rational structure of the world and its objects. We open a bracket: the failure of the enterprise of knowing that the skeptic declares is not very credible because, even to sanction the failure, reason is needed and even filing for bankruptcy is a rational declaration.

In reference to the human subject, reason is the ability to implement strategies for understanding the world and ourselves, it uses arguments. It therefore knows how to orient itself between different hypotheses and scenarios, thanks also to the contribution of emotions and the body with whose needs it intertwines its path.

Between objective reason and subjective reason there is a fundamental asymmetry that inevitably derives from their different structure, but "pure" objectivity, that is to say objectivity without a subject that catches it, is unattainable, as idealism has well noticed: there is no object except for a subject who knows it (Calogero, 1927), while the reverse (no subject without object) is at least an unexplored field. The dynamics between subject and object is a continuous flow of mutual transformations and the attempt to know the object regardless its interplay with the subject is a rather dangerous myth, which drives us to seek the unobtainable.

Today the vital trace of idealism can be found in the "sourcentist" positions, (Maturana & Varela, 1980) which affirm that the subject makes the world arise in the act and in the way of knowing. Unlike the idealist, the "sourcentist" is not obliged to conclude for the mental existence of the world, but he can argue that reality is a domain of reference external to the subject, the identification and description of which pertains to the subject, to its objectives and capabilities.

## 3.2 From an Epistemological Point of View Reason Can Be Implicit and Explicit

While the "explicit" reason is well known to philosophers and also to the on-going experience of human subjectivity, which always talks to and with itself in ways that can be related, remembered and even traced, a meaningful part of the rational activity that does not reach the conscience has been largely ignored or denied. We refer to the "implicit" reason, whose effectiveness is revealed in what we do, that governs our behaviors and also acts without our knowledge, with motivations and orientations not being necessarily present to the consciousness. Aristotle calls it *noesis* and attributes to it a regulating force much wider than that recognized to *dianoetic* reason, which translates the dictation of *noesis* into the propositional discourse, subjected to the constraints of syntax and logic.

Vitiello (2019) expresses a not dissimilar concept in the field of theoretical physics when he states that the microscopic phenomena that support macroscopic manifestations are intrinsically opaque to knowledge: reason admits them and identifies them at least in part, even though it cannot provide their description, if not indirectly.

It is easier to recognize the explicit reason, which exercises its action in a traceable way. Its path can be followed in the areas in which it is active and does not require the use of abduction to pass from the visible to the invisible: it can be expressed through the classic tools of logic, which organize knowledge data in a formal structure.

## *3.3 From the Point of View of the Knowing Subject Reason Can Be Conscious and Unconscious*

The Freudian discovery of unconscious processes has not weakened reason by handing it over to a dark and unknown region, but it has extended its breadth: even the unconscious with its proper dynamics, traceable only through interpretation of universally shared phenomena—the well-known dreams, slip acts, discomforts and psychic pathologies—can be understood thanks to the guidance of reason.

A guideline which in this case prepares and uses observation, access and control tools built in order to investigate a field that can be intellectually grasped only indirectly. Also in this domain there is no lack of anchoring to the empirical, which is the therapeutic capacity of the different psychoanalytic approach to interact with the hidden world of the patient, bringing about improvements or remission of pathologies at least in the cases where they are successful.

It would be interesting and worthy of a discussion the problem about the violation of the principle of contradiction observed in several interpersonal conflicts: I can love and hate the same person at the same time, I wish to meet someone and I can't stand to see him. One could suggest the hypothesis that the violation of the principle of contradiction experienced at an unconscious level is at the origin of mental illness.

To verify this hypothesis it would be necessary to build an interdisciplinary research project, with the collaboration of philosophers and psychoanalysts. If such conjecture were proved, we would confirm furthermore the divergence between logic and rationality: the contradiction is a violation of logic—which only possesses tools for reporting the violation, but would not know how to exploit and understand it. But it does not violate rationality that even in this case would pervade any human action: pathology expresses and denounces an "impossible" experience because it is contradictory in the only language available to it, that of suffering. In fact, as it is well known, the unconscious does not have access to the discursive and propositional language of consciousness and is expressed through images, metaphors, and symptoms.

According to Libet (2004), a neuroscientist well known for his experiments in the field of brain and the mental *timing*, alongside the conscious mind, there is the unconscious mind, which governs many human actions, including the complex and intertwined operations of physiology, aimed at the organism's survival. It solves many problems, even theoretical ones, silently, providing the conscience with the result of its work. We are talking about an unconscious mind which, Dupré might add (1995, 2003), in an imaginary remote dialogue with the sciences, is action-oriented ("I press the brake so that I don't kill a cyclist crossing the roadway"), introjects cultural imperatives and models.

A further insight about this topic comes from Stern's (2004) distinction between unconscious and nonconscious mind, which provides a more comprehensive set of the dynamics of the Ego.

From these and other researches, the profound unity of the human emerges with even more solid evidence: psychic states influence choices, reason is imbued with endogenous and exogenous relationships, the unconscious mind interacts with the conscious mind. The weft of the human is woven by intertwining all these threads and each of us expresses and lets himself/herself be seen in its intertwining. Hence we face a fundamental challenge that consists of refuting the idea that mind primarily and exclusively follows logical rules. To provide some valuable motivations to support this criticism, it is useful to begin by exploring the possibility of a more extensive feature of reasoning. New lenses are required to adequately support this task.

## *3.4 From a Logical Point of View Reason Can Be Abductive (Creative) and Deductive (Tautological)*

Reason, in particular in its explicit value, applies different inferential modalities to achieve the cognitive objectives it has set for itself. In a very general way, inference consists of the necessary or at least possibly provisional connection between a proposition that is deemed true and a subsequent and dependent one. The inferential modalities traditionally studied in logic and epistemology are deduction and induction; the deduction that is drawn from true premises leads to necessarily true conclusions, while induction obtains, at most, a cognitive result of high probability as it generates laws or rules starting from a limited number of cases. Due to the necessity and certainty of its conclusions, deduction remains the leading tool for logic, with the well-known limit of producing tautologies.

Yet human knowledge is not only tautological or anchored to observational data: on the contrary it is often innovative and capable of progress (very different from accumulation). On closer inspection, human beings mostly use another inferential modality—abduction—the results of which are not certain, and often not even probable. Abduction is widely and successfully used not only in daily practice, but also in many areas of the greatest relevance, such as medical diagnoses, police investigations. More widely this inferential modality is applied in all occasions in which a "new", "creative", "unexpected" theoretical or practical result is reached; this is the case with scientific theories. It is due to systemic thinking that we have brought attention back to this kind of inference, already known to Aristotle 1924, (*Prior Analytics*: 2, 26, 69 a, 20–38) and Peirce (1931–1958), and which is now beginning to enter the logic manuals (Frixione, 2007). But what is abduction? (Urbani Ulivi, 2016).

Just to have a reference platform, we suggest two definitions for "abduction" and for "creativity".

> By abduction we mean an inference that, by operating in an incomplete information context, over-determines the available data and identifies either a universal hypothesis (a law or a theory) or even a particular object in response to the question posed by the investigation.

> By creativity we can mean, at least in a useful approximation in this context, the outcome of a procedure that cannot be formalized through given rules.

Abduction can be of various types, but at this point of the paper we aim to underline the *creative abduction* as the discovery of the hypothesis that organizes the entire cognitive landscape in a new way and that cannot be prescribed through a formal procedure; it follows that it is impossible to bring creative abduction back into the context of formal logic, for which abduction is nothing more than an incorrigible anomaly. Obviously, the investigation on abduction remains open to further philosophical investigations, which require to widen the context of reference. In other terms, we should answer a main question that can be formulated as such:

> How knowledge must be re-thought so that it can also include abduction, creativity, information incompleteness, variable contexts, non-deductive and not even inductive inferences?

There is but one answer: once we recognize that the formal rules of logic characterize a part of the knowledge activity, but do not exhaust it or even complete it, we must admit that thought processes draw on many and different resources that go beyond any formal procedures. They root knowledge in a real world and, first and foremost, in a subject embodied in the personal and affective relationships: all these aspects are part of the individual's personal history, in the social, political, religious bonds, which enter into the argumentative procedures, orient them, support them, being only partially recognized and explicitly recognizable.

At this point of the investigation we have reached the intermediate result to focus an enlarged view of reason, embedded in structural, dynamic relationships with the environment of which human beings are included. This path, by shaking all the current paradigms, invite us to turn on a systemic perspective.

### 3.5 From an Anthropological Point of View Reason Is a Second Level Property of the Human System

We generally say, through the lenses of the system thinking, that a property is a "second level property" when it depends on the entire constitution of an entity. In the case of humans, reason operates taking into account all those factors whose intertwining qualifies and identifies each subject.

We talk about body, emotions, personal history, preferences, moral, family, social constraints, freedom, and much more that is typical of the human being. Reductionist simplifications cannot be followed: such a thesis cannot be sustained, at large, because offering a simplified explanation of a multi-faceted phenomenon it misrepresents the phenomenon as a whole, only describing some of its constituents.

To provide some intuitive motivation for this remark, it could be useful to recall some examples. "It is rational to eat if you are hungry, to drink if you are thirsty"—this is the deceptively persuasive example used by many philosophers to suggest the universality and sharing of the criteria of rationality, but it is an example, precisely, deceptive: I'm hungry, there is food, it is not always rational to eat it. Indeed, it is rational that I do not eat it if I'm waiting for the guests for whom the food has been prepared (cultural constraints) if I want to lose weight (wishes to which hunger is subordinated), if by eating that food I steal it from my child if the ongoing famine does not allow further supplies, etc.

Human action, of any type, from the most corporeal to the emotional, sensitive, deliberative ones, is as rational as it is capable of prefiguring different scenarios with the results of the various actions undertaken. And it is as rational as it is able to choose which scenario to give course, but reason does not oblige us to make a specific choice, nor it does make the same scenarios available for everyone. It is not a universal criterion, it is a principle that activates scenarios and strategies of orientation—practical and theoretical—that we use to understand different situations and to behave in different circumstances.

The "ways", in the sense of paths, of reason are neither homogeneous nor even equal, despite having in common the achievement of a goal. The battles are different, but while you need an enemy to make a battle, for there to be "reason" you need a purpose. Purposes are not rational—nor irrational—they are pursued with rationality.

Some might prefer to the term "way"—which remains very approximate and vague, similar in this to the term "manner"—the term "procedure", more precise. However, we suspect that, by procedure, we inevitably mean a coordinated series of passages formally codified by logic. Of course, reason is also this: logical-formal procedure, but it does not only reduce to this, being a complex texture of a variety of threads. It is also choice, decision, preference, appreciation of some aspects, carelessness of others, it is tears, it is laughs. For none of these traits we would ever use the term "procedure", while we could speak of "mode", recalling the Cartesian sense of mode as a "variable or transient quality", or Aristotelian pluralism (the modes of being), to arrive at the grandiose Spinozian construction, which saw in the modes the necessary "affections" of the substance.

To the "mode", even if differently declined in different eras and authors, the plurality of expressions appears intrinsic, very suitable to reflect the operative plurality of reason, and its unpredictable, different, surprising, yet rational outcomes.

## *3.6 From a Topological Point of View Reason Can Be Local and General*

Once the dangerous illusion of a universal reasoning criterion that governs every cognitive domain in the same way has been abandoned and the simple recognition of a plurality of philosophical visions *à la* Rescher (1985) is considered insufficient, one cannot but accept and put forward a pluralist perspective (very different from perspectivism). Declined on a topological horizon, pluralism affirms that reason governs each domain of knowledge with local criteria, within which the conditions of acceptance or rejection of hypotheses and theories are established, and that in addition to local criteria there are validity criteria that going beyond the local ones we can perhaps call "general" (we speak of "general" with some caution, because it is a generality connected to the different "localities"). It should be recognized that the general criteria by increasing the extension, fatally decrease the intension, whereas the local ones by increasing the intension lose in extension.

In this perspective, general and local criteria interact and are linked by relationships of interaction and interference; they are mutually open to changes and developments that transform them over time. To the various characteristics of reason, another one should be added: it is processuality, which makes it flexible and adaptable to the different needs of its presence in history.

## *3.7 Pluralism Versus Monism*

There are no formal and universal criteria for rationality, valid in any circumstance, for any subject, in any environment. Reason adopts different validity criteria according to the domains intended. What is rational—or inspired by reason—here and now is not exportable—or it is not always exportable, nor should it be expected to be—elsewhere and in another time. Therefore monism and universalism of criterial reason and logic should be replaced with a pluralistic view of reason.

Rescher in his *The Strife of Systems* (1985), observing the coexistence of philosophical systems on the scenario of thought, each of which claims to be universally valid, to avoid the easy skeptical drift, proposes the "pluralism of orientations", which affirms that in philosophy coexist acceptable alternatives, although they may have very different merits, therefore many different orientations in terms of setting and results are admissible on the philosophical scenario. Rescher's position, by recognizing that there are different competing philosophies on the scenario of thought, does express a fact, but does not explain why these differences exist. In order to explain this wide range of perspectives (hence perspectivism), reason should no longer be considered as an universal tool that exercises its activity according to universal procedures and rules, but in an authentically pluralistic way, as an activity of understanding that changes according to the historical moment, of the circumstances, of the problematic area to which it refers, of the objectives to which it is addressed, of the implicit but powerfully influential assumptions that it adopts, of the own and individual sensibilities of each philosopher, of his moral structure, of his aesthetic sensibility, and of much even more.

Philosophical systems are not ahistorical, disembodied, absolute, they do not express a criterial reason independent of the circumstances, but they represent the effort with which each generation and each individual tries to understand the world with the tools of knowledge available at a certain time and with different capacity and sensitivity of each. A pluralistic position—very different from the obvious admission that different systems coexist and struggle, *à la* Rescher—explains the plurality of positions by introducing not only, but also logical reasons as forces that structure and define different positions.

## 4 What About "Irrational" and "Unreasonable"?

"Irrational" is not a term predicable of "man". Losing one's reason—understood in the extended sense proposed here—means losing humanity. The subjective reason may be missing, but the unconscious reason or even only the objective reason will remain to structure the human, even if only in the organization of corporeality. The human subject can carry out actions whose *ratio* remains opaque to the observer and also to the one who performs them, but the embodied, hidden, inspiring *ratio* is always there: the mentally ill person, the immoral, the criminal, behave following questionable criteria, or not shared by others, but they follow reason. Their reason. Which can be misused, superficially, erroneously, counterproductive, but it is still followed. Of course, there are "irrational" or "unreasonable" actions, but they are precisely single actions that do not undermine the general structure of reason characteristic of human behavior.

This passage is highly relevant because it is expected to reshape the dominant paradigm about the human agency and to have consequences in several fields of the human sciences, from psychology and psychiatry to law. It's interesting to underline that, regarding reason, contemporary researches in psychotherapy are moving towards the same path traced by the systemic thinking approach. We could not understand why therapy brings about a change in the psychological discomfort if we would not consider that *sui generis* rationality. The possibility of any psychological treatment passes by the capacity of the therapist to become acquainted with the mental world of the patient, both cognitive and emotional: therapeutic approaches like the EMDR (Eye Movement Desensitization and Reprocessing), which works with the opaque dimension of the mind, proves that any presumed "irrational" content is highly "rational". Worth noting that not the logic but a processual approach—the EMDR counts 8 different phases—both cognitive, emotional and sensitive (the therapist's hand tapping) is the very key to enter the intimate ground of the patient. A logic approach is taken on by the two actors of the process, the patient and the therapist, only in the "assessment" (phase 3) and, at the end of the therapeutic path, in the "closure" (7) and "evaluation" step (8) (Fisogni & Fisogni, 2020).

Furthermore, in the last two decades the phenomenon of global terrorism has deeply put in question what "reason" is and how reason works in evildoing.

There is a wide consensus among scholars that global terrorists are not irrational, although their acts seem to be completely out of reason. It is questioned whether they perform their attacks according to a limited critical capacity, due to the ideological desensitization, or their acts are the result, on the cognitive ground, of a lack of empathy (Fisogni, 2010) that is amplified by the turn from the offline to the digital domain. Finally, exploring the opaque dimension of the reason also means to re-open the ethical debate about the major phenomena of evildoing, like the death of millions of innocent persons in the Nazi's extermination camps.

### *4.1 Skepticism Towards Reason Defeats itself*

The procedures and outcomes of reason are the thread that guides us into the investigation of reality: every attempt at understanding is radically antisceptic; the outcome can be uncertain and provisional—or even, as for the skeptic, negative—but trying to understand is an activity imbued with optimism, animated by a fundamental trust in rational activity and by the constant and continuous use of reason.

We are well aware of the fact that many authoritative voices of philosophical thought have risen against this statement (Pascal, 1901; Kierkegaard, 1944; Schopenhauer, 2014; to a certain extent also Hegel, 1807; Bergson, 1907). These authors deserve the credit for having grasped and denounced the limits of reason understood only as logic, which is criticized for the impossibility of understanding those areas in which logic has no place to proceed. The defect of these "irrationalist" positions has been to adhere to a reductive and limited concept of reason, which has not been able to see how much reason goes beyond logic and impregnates every human activity. Their denunciation is not valid against the expanded and pluralistic concept of reason that we have proposed, while it remains very effective in demonstrating the limits of a restricted reason within the bed of logic.

## 5 Conclusion, Knowingly Open and Provisional

This inquiry took as its starting point the question: why do we return once again to a subject so widely discussed, plowed up and debated? The answer could be this one: for an unshakable confidence in the progress of knowledge. We think that although knowledge does not progress in a linear way or even by accumulation, it is necessary to exploit the undoubted cognitive advantages offered by the contemporary world and by the tools that have been discovered, developed and made available by it. We are thinking in particular of systemic thinking and its theoretical strength that is starting to be exploited. It has been proved capable of expanding our capacity for understanding in several areas, and of replacing worn-out tools that have frequently proved insufficient or unsuitable to solve many theoretical or practical problems.

We find different terms in Greek: *logos*, *nous*, *dianoia*, and non-corresponding other terms in Latin: *ratio*, *intuitus, mens, intellectus*. All these terms belong to a conceptual family in which there are notable similarities and divergences. For the differences, a brief historical investigation is sufficient; but where do they converge? What is the platform of reference which is inclusive of the different expressions of reason? As we have noticed before, reason is not reducible to *dianoia*, to any judgment argued in a logical form, but it is also not reducible to *nous,* nor to the rich and deep *logos.*

Every *dianoia* takes place starting from a *noetic* basis, from a *nous* understood as the ability to intellectually grasp the elements that will subsequently constitute the judgment: that too is the business of reason. It is *dianoia*, the connection of separate parts, but it cannot be reduced just to *dianoia*; it is explicit and conscious, but its foundations are rooted in the unconscious mind and bodily activity. It is subjective, but it acts by tracing the reason that objectively structures the world of phenomena.

Reason therefore is a principle of organization and order of both reality and knowledge. We talk about "principle" in the sense of "what regulates"; for example it regulates the dynamic of a process and allows it to be distinguished from the environment in which it is immersed, while the order is mainly the result of an organization of objects (both concrete, theoretical or even mental) between which relationships are established. It follows rules, laws, prescriptions, constraints. It results from negentropy, from symmetry breaking, from self-organization.

Reason has many aspects and shows different faces depending on the focus of each investigation: it is, and can be, subjective and objective, conscious and unconscious, explicit and implicit, pragmatic and theoretical, argumentative and apprehensive, and much more.

It is necessary to abandon any illusory ideals of reduction, it is necessary to admit more than what contemporary logicisms have accustomed us to recognize: reason is not only the procedure that guarantees scientific knowledge, but it is, within the sciences, also that ability to formulate hypotheses that precedes the control procedures.

In a word, it is creative.

It favors vagueness and incompleteness: it renounces misleading claims of accuracy, completeness, which have remained unrealized announcements and proclamations. It investigates the asymmetries that govern the transformation of *chaos*—where everything is interchangeable because it lacks identity characters—into *cosmos*, rich of order and of orderly processes, distinguishable because they are equipped with emergencies that make them identifiable. If we want to use a metaphor that makes our idea clearer, the reason is similar to the road for those who walk. To make any walk possible, there must be a road, and it can be flat or uphill, well defined or barely outlined, smooth or coarsely cobbled. It must be accepted and recognized that not all roads are the same—there are 10-lane highways and mountain paths—and one cannot give a universal description, which is suitable and appropriate for all roads. What is good for everyone is minimal: a road establishes a limit with respect to the environment (external limit) and allows different movements within it: the pace can be fast, slow, depending on who takes it, why it is accomplishes.

We reached the end of the paper, not of the investigation, which is expected to be developed through further interdisciplinary suggestions and criticism. The scope of our proposal is intended to be a preliminary, however well-argued step addressed to rethinking what reason is, what it does, how it acts, what results it can veritably achieve. In search of a more comprehensive understanding of such a challenging subject matter, we are perfectly aware that the shifts of reason are different and often unexpected: they must be followed carefully, with patience and with optimism.

## References

Aristotle. (1924). Metaphysics. In *Edited with introduction, commentary and translation by W.D. Ross*. Clarendon Press.

Bergson, H. (1907). *L'Evolution Créatrice*. Alcan.

Blackburn, S. (Ed.). (1994). *Dictionary of philosophy*. Oxford University Press.

Calogero, G. (1927). *I Fondamenti della Logica Aristotelica*. Le Monnier.

Damasio, A. (1994). *Descartes's error. Emotion, reason and the human brain*. A. Grosset-Putnam Books.

Damasio, A. (2010). *Self comes to mind. Constructing the conscious brain*. Pantheon Books-Random House.

Descartes, R. (1641). *Meditationes de Prima Philosophia*. Soly.

Devlin, K. (1997). *Goodbye, Descartes: The end of logic and the search for a new cosmology of the mind*. Wiley.

Dupré, J. (1995). *The disorder of things. Metaphysical foundations of the disunity of science*. Harvard University Press.

Dupré, J. (2003). *Human nature and the limits of science*. Oxford University Press.

Fisogni, P. (2010). Terrorists: Analogies and differences with mental diseases. A phenomenological-metaphysical perspective. *Rivista di Psichiatria, 45*(3), 145–153.

Fisogni, P., & Fisogni, A. (2020). Temporality and the traumatic memory treatment. *Clinical Research in Psychology, 3*(1), 1–6.

Frixione, M. (2007). *Come ragioniamo*. Laterza.

Hegel, G. W. F. (1807). *Phaenomenologie des Geistes*. Goebbards.

Kierkegaard, S. (1944). In V. Eremita (Ed.), *Either-or: A fragment of life*. Princeton University Press.

Libet, B. (2004). *Mind time. The temporal factor in consciousness*. Harvard University Press.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Reidel.

Minsky, M. (1986). *The society of mind*. Simon & Schuster.

Nozick, R. (1993). *The nature of rationality*. Princeton University Press.

Pascal, B. (1901). *Thoughts*. Bell.

Peirce, C. S. (1931–1958). Upon logical comprehension and extension. In *The Collected Papers of Charles Sanders Peirce* (2643 ff). Harvard University press.

Piattelli Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. John Wiley & Sons.

Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.

Rescher, N. (1985). *The strife of systems. An essay on the grounds and implications of philosophical diversity*. University of Pittsburgh Press.

Rescher, N. (1988). *Rationality: A philosophical inquiry into the nature and the rationale of reason*. Clarendon Press.

Schopenhauer, A. (2014). *The world as will and representation*. Cambridge University Press.

Simon, H. (1983). *Reason in human affairs*. Stanford University Press.

Stern, D. N. (2004). *The present moment in psychotherapy and everyday life*. W.W. Norton.

Stich, S. P. (1991). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. MIT Press.

Turing, A. (1992). In D. C. Ince (Ed.), *Mechanical intelligence*. North Holland.

Urbani Ulivi, L. (2016). L'abduzione come Momento della Scoperta in Contesti di Realtà. Abduction is the inference that discovers a solution in problematic contexts. *Cassazione Penale, 56*(5), 2240–2251.

Vitiello, G. (2019). The world opacity and knowledge. In L. Urbani Ulivi (Ed.), *The systemic turn in human and natural sciences* (pp. 41–51). Springer.

Winograd, T. A., & Flores, F. (1986). *Understanding computers and cognition: A new Foundation for Design*. Ablex Publ Corp.

## *Bibliography*

Readers do not find any bibliography, and this choice ought to be at least clarified. We gave up on it because a bibliography, while not complete or at least reliable, would have been probably manageable with difficulty because of its immense extension. We could adopt selective criteria, one could say; yes, we could, however even this hypothesis was discarded because any selected criterion would have been tailored to personal interests and sensibilities. We therefore leave to the reader to pick up his/her personal choice of reference texts. We would like to suggest—in particular to the less experienced—to let themselves be guided by the entries "reason", "intellect", "intuition", "argumentation", "inference", "logic", "mind", of the Stanford Encyclopedia of Philosophy and the Oxford Dictionary of Philosophy.

# Neural Networks and Many-Body Systems

**Giuseppe Vitiello**

**Abstract** I review one of the papers by Eliano Pessa and myself where we model neural networks as many-body systems. The mathematical formalism underlying the model is here presented. The information input induces the spontaneous breakdown of the $SU(2)$ symmetry of the units of the neural net. The order parameter describes the collective correlation modes propagating over the net and represents the code of the recorded information. The dissipative character of the dynamics allows large memory capacity. Free energy minimization leads to the sigmoid activation function for the neural net units. Entropy and free energy play a crucial role in characterizing the net functional properties.

## 1 Introduction

It is very sad to write a contribution for a book in memory of Eliano, a companion on the road in the search of some thing that could satisfy our common passionate "desire to know". The memory of the excitement created by the fact that such "thing", appearing first as an intuition, could be then expressed in mathematical language, makes it all the more sad, since emotions cannot be cut, left without who shared them with you. With Eliano there was a tacit understanding, the silent friendship that binds the seekers in a common adventure. Perhaps the same one that united him with his teammates in climbing the mountains he loved so much.

I then thought that summarizing and commenting here one of the articles (Pessa & Vitiello, 1999) we wrote together could somehow alleviate my sadness.

In few words, the underlying attempt of our search was complementing the "atomistic" view prevailing in science in those years, and still dominating today, with the understanding of the collective dynamics responsible of the behavior of the elementary components as a unified system. *Naturalism*, namely the list of the

---

G. Vitiello (✉)
Dipartimento di Fisica, Università di Salerno, Salerno, Italy
e-mail: vitiello@sa.infn.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_14

elementary components and all their properties in all possible details, is *necessary*, but yet *not sufficient* to knowledge. Naturalism is not yet *science*. One needs in fact to know also the *dynamics* ruling the elementary components so to reach the understanding of their collective behavior.

Eliano was an expert in neural networks and our effort was to model the states of such nets in terms of collective modes, a description which we thought to be possible to formulate in terms of quantum field theory (QFT).

Actually, from one side Haken and collaborators (Haken, 1991) proposed to model artificial neural nets by studying collective modes in classical and quasi-classical formalism. On the other hand, Ricciardi and Umezawa (1967) pointed to the possibility to describe neuronal activity in natural brains by use of collective modes in QFT. Their model and subsequent developments (Stuart et al., 1978, 1979; Sivakami & Srinivasan, 1983) led to the inclusion of dissipative dynamics in the study of brain modeling (Vitiello, 1995, 2001, 2004a; Freeman & Vitiello, 2006). These studies were realized also on the basis of the role played by coherent dynamics, such as the one of holograms in laser physics (Pribram, 1971, 1991; Jibu & Yasue, 1995; Jibu et al., 1996). Of course, an enormous amount of successful work in the study of (artificial) neural nets was already done and was still in full development in the frame of statistical mechanics and spin glass theory (Mézard et al., 1993; Amit, 1989). Moreover, the studies of computational tools based on quantum mechanics (quantum computing) were also rapidly developing (Nielsen & Chuang, 2000).

The results of Pessa and Vitiello (1999) are summarized in Sect. 2. The proof of the dynamical formation of long range correlations and of the coherent condensation of the associated Nambu-Goldstone (NG) quanta is presented in Sect. 3. Some mathematical details are given in the Appendix. The dissipative structure of the dynamics is discussed in Sect. 4. Entropy and free energy minimization and the relation with the sigmoid activation function are presented in Sect. 5. Section 6 is devoted to the conclusions.

## 2 Neural Nets and Field Description

In Pessa and Vitiello (1999) the activity of a neural unit is assumed to be characterized by the amplitude of the emitted pulse and by its phase determined by the emission time. Thus, in a set of $N$ neural units at the space-time sites $x_n \equiv (\mathbf{x_n}, t_n)$, with $n = 1, 2, ..N$, each one of them may be described by an $SU(2)$ complex doublet field

$$\psi(x_n) = \begin{pmatrix} \psi_u(x_n) \\ \psi_d(x_n) \end{pmatrix} \quad . \tag{1}$$

The suffixes $u$ and $d$ of the complex components $\psi_u(x_n)$ and $\psi_d(x_n)$ denote the inner degrees of freedom corresponding to the *on* (1) and *off* (0) states of the unit, respectively, analogous to spin degrees of freedom in fermion fields. The state of each unit depends on its interaction with the other units, which, at this stage of our discussion, does not need to be specified. In full generality, it is only assumed that such an interaction is symmetric under $SU(2)$ transformations of the $\psi(x_n)$ fields in the $u - d$ space.

In our picture, the *functional state* of the net, also denoted as its *macroscopic state*, does not correspond to a single, uniquely defined *microscopic* configuration of values $u$ and $d$ at the sites $x_n$. Therefore, many distinct microscopic configurations may in principle correspond to the same macroscopic state of the net. Which, said differently, means that one specific macroscopic net state does not depend specifically on the activity of one single unit, neither on the *fluctuations* of each unit between its $u$ and $d$ states. We call them microscopic fluctuations and refer to such a property of the net as its *plasticity* (rigidity denotes the opposite case of strict dependence of the net functional activity on the definite state of each of its $N$ units). Although our discussion refers to artificial net, the assumption of the plasticity property is suggested by the observation of the natural brain, whose functional activity does not depend on the activity of each single neuron (Ricciardi & Umezawa, 1967; Freeman, 2000).

The independence from the (microscopic) fluctuations allows us to associate to the macroscopic state of the net a “classical” field, called “order parameter” in analogy to condensed matter physics, where the order parameter is independent of quantum fluctuations. The word ‘classical’ is used to mean, indeed, ‘independent of microscopic fluctuations’.

The order parameter is related to the symmetry properties of the interactions of the units (Umezawa et al., 1982; Takahashi & Umezawa, 1975; Umezawa, 1993; Itzykson & Zuber, 1980; Anderson, 1984; Blasone et al., 2011a) and acts as a *code* labeling the least energy state (the vacuum or ground state) of the net (Ricciardi & Umezawa, 1967; Vitiello, 1995), i.e. the information content of the net.

In Sect. 6, I will comment on the use of the quantum, or quantum-like, formalism in the description of neural nets, also in connection with ’t Hooft works on dissipative systems appearing as quantum systems (’t Hooft, 1999; Blasone et al., 2001b).

By using the same notations of Pessa and Vitiello (1999), $\mathcal{M}$ denotes the order parameter of the net. It is defined by

$$\mathcal{M} \equiv \frac{1}{2}|(N_u - N_d)| \, , \tag{2}$$

with very large $N$ (infinite in the infinite volume limit):

$$N = N_u + N_d \, . \tag{3}$$

$N_u$ and $N_d$ denote the number of units *on* and *off*, respectively.

The state characterized by $\mathcal{M} = 0$ is said the “normal” state and it is assumed to be void of information content. States with $\mathcal{M} > 0$ are “information states”, with different non-vanishing values of $\mathcal{M}$ associated to different information contents. These states are also called "memory" states. Since for any $\mathcal{M}$ we have a zero-energy state, infinitely many degenerate (vacuum) states exist in the continuum or infinite volume limit, each one corresponding to a different information content (a different value of the memory code $\mathcal{M}$). In the following sections, it is shown that there is no unitary transformation connecting these vacua. They are unitarily inequivalent states.

As said above, the dynamics controlling the interactions among the units is assumed to be invariant under $SU(2)$ transformations. The normal state, however, is not invariant under such transformations. They change it into a state with a given non-zero $\mathcal{M}$. One then says that spontaneous symmetry breakdown (SBS) occurs. The external input carrying information acts as a trigger inducing SBS. The coupling of the net with the environment thus plays a crucial role in selecting a specific vacuum state among the many unitarily inequivalent vacua available to the net. In Sect. 3 it is shown that SBS generates long range correlations among the units of the net.

A given value of $\mathcal{M}$, i.e. of the difference $|(N_u - N_d)|$, does not fixes the specific value of $N_u$, or of $N_d$, which means that the "observed" value of $\mathcal{M}$ is compatible with many distributions of the states of the units determined by the *microscopic* configurations of the $\psi(x_n)$ fields. In this sense $\mathcal{M}$ appears to be a *macroscopic* variable.

The conclusion is that, as observed above, from the net functional activity standpoint, it is not relevant which one is the state, $u$ or $d$, of a specific unit. Indeed, units may be excited or de-excited by pulses (signals) traveling over the net. On the other hand, any observation of the state of the units may interfere with the traveling pulse. This means that the states of the units should be considered as non-observable, unless to perturb them in a way to drastically deform the *dynamical* distribution of pulses on the net and thus its functionality. Only the *output* of the dynamics of the units is observable, namely the state of the net at a time $t$ at which the (quasi-)equilibrium state has been generated; it is called the *asymptotic* state at time $t_n \to \pm\infty$ for each $n$. Corresponding to the $\psi(x_n)$ fields, complex doublet fields $\phi(x_n)$ may be introduced describing the units in their asymptotic state.

In this description the fields $\psi(x_n)$ (and $\phi(x_n)$) can be assumed to be quantum, or quantum-like, fields. The neural units, however, are not considered to be quantum objects, rather the fields describing their dynamical behavior are assumed to admit formal quantum-like description.

In the following section, the memory states will be recognized to be coherent states (Klauder & Sudarshan, 1968) arising from the condensation of quanta of long range correlations among the units and such a coherence is the source of the independence from microscopic fluctuations of the order parameter $\mathcal{M}$.

## 3 Generation of Long Range Correlations

Let me now present some of the formal aspects behind the description presented in the previous section (and in Pessa and Vitiello (1999)). This formalism is similar to the one describing ferromagnets with localized spin components, where the symmetry is the $SU(2)$ continuous rotation group in the spin space (Umezawa et al., 1982; Blasone et al., 2011a; Shah et al., 1974).

The interaction among the $\psi(x_n)$ fields associated to the units of the net is given in terms of a Lagrangian whose explicit form is not necessary to specify. It is only assumed that it is invariant under the $SU(2)$ group of continuous rotations in the $(u-d)$-space.

In the following, by adopting the same notation used in Umezawa et al. (1982), Blasone et al. (2011a), Shah et al. (1974), let $S^{(i)}(x_n)$, $i = 1, 2, 3$, $n = 1, 2, ..N$, denote the generators whose $SU(2)$ algebra is:

$$[S^{(i)}(x_n), S^{(j)}(x_l)] = i\epsilon_{ijk} S^{(k)}(x_n)\delta_{n,l}. \tag{4}$$

Their explicit form in terms of the anticommuting fields $\psi(x_n)$ is given for example by $S^{(i)}_{\psi}(x_n) = \psi^{\dagger}(x_n)(\sigma_i/2)\psi(x_n)$, with $\sigma_i$ the Pauli matrices. Our discussion and conclusions will be however independent of the specific form of $S^{(i)}(x_n)$.

The (total) $SU(2)$ generators are

$$S^{(i)} = \sum\nolimits_n S^{(i)}(x_n), \quad i = 1, 2, 3, \tag{5}$$

$$[S^{(i)}, S^{(j)}] = i\epsilon_{ijk} S^{(k)}. \tag{6}$$

Under $SU(2)$ $\psi(x_n)$ transforms as

$$\psi(x_n) \to \psi^{'}(x_n) = \exp(i\theta_i\lambda_i)\psi(x_n), \qquad i = 1, 2, 3, \tag{7}$$

with $\lambda_i = \sigma_i/2$ and $\theta_i$ a triplet of continuous group parameters (the rotation angles in the $(u-d)$-space). For $\theta_i$ infinitesimal $S^{(i)}(x_n)$ transforms as as

$$S^{(i)}(x_n) \to S^{(i)'}(x_n) = S^{(i)}(x_n) - \theta_j\epsilon_{ijk} S^{(k)}(x_n). \tag{8}$$

Invariance of the Lagrangian implies: $L[\psi(x)] = L[\psi^{'}(x)]$; the ground state $|0>$ is however assumed to be not invariant under the full $SU(2)$ group but only under the subgroup $U(1)$ of rotations around the 3rd axis in the $(u-d)$-space. We have thus SBS. This last assumption describes the fact that an external stimulus, due to an information source and acting on the neural net, may modify the ground state of the neural net by producing a non-zero expectation value of the $S^{(3)}$ generator:

$$< 0|S^{(3)}|0 >\equiv \mathcal{M}, \tag{9}$$

where $\mathcal{M}$ is indeed the "memory" order parameter introduced in the previous section. The information carried by the input is thus recorded in the state $|0 >$ and labeled by $\mathcal{M}$. This is made explicit by writing: $|0 >\equiv |0 >_{\mathcal{M}}$.

According to a general, experimentally well verified, theorem of QFT (the Goldstone theorem) (Umezawa et al., 1982; Itzykson & Zuber, 1980; Blasone et al., 2011a), the SBS implies the dynamical formation of long range correlations among the system elementary constituents (the units of the net in our case). The boson quanta associated to such correlation waves spanning the system are called Nambu-Goldstone (NG) quanta and denoted $B(x_n)$ and $B^\dagger(x_n)$, maintaining the notation of Blasone et al. (2011a), Shah et al. (1974). These boson quanta are massless since they span the whole system and, at their zero momentum, do not add energy to the ground state $|0 >_{\mathcal{M}}$. Their associate correlation waves propagate over the net without destructive interference, namely in a *coherent* way. In other words, the *coherent condensation* of NG boson modes (the $B$-modes) in $|0 >_{\mathcal{M}}$ is obtained.

It is useful for our discussion to recall briefly the proof of the Goldstone theorem. I will adopt the model independent functional integration formalism (see Umezawa et al. (1982), Blasone et al. (2011a), Shah et al. (1974)).

In a standard notation, the Green's function generating functional is

$$W[J, j, n] = \frac{1}{N}\int [d\psi][d\psi^\dagger]\exp\{i\sum\nolimits_n \int dt L[\psi(x_n)]+J(x_n)\psi(x_n) \tag{10}$$

$$+\psi^\dagger(x_n)J(x_n) + j^\dagger(x_n)S^{(-)}(x_n)+S^{(+)}(x_n)j(x_n) + S^{(3)}(x_n)n(x_n) - i\epsilon S^{(3)}(x_n)\},$$

where $N$ is a convenient normalization factor, $J$, $j$ and $n$ denote the source fields, $S^{(\pm)}(x_n) = S^{(1)}(x_n) \pm i S^{(2)}(x_n)$. The $\epsilon$-term represents the coupling with the environment. It describes the informational input inducing SBS. Its action is limited in time and does not change the internal dynamics of the units described by the Lagrangian $L$. It affects however the state of the system. The limit $\epsilon \to 0$ has thus to be taken at the end of the computation.

Consider now the Ward-Takahashi relation (Shah et al., 1974) derived from (10):

$$< S^{(3)}(x_m) >_\epsilon = \epsilon\sum\nolimits_n \int dt < S^{(i)}(x_n)S^{(i)}(x_m) >_\epsilon , \quad i = 1, 2. \tag{11}$$

It shows that, in order to obtain a non-zero expectation value for $< S^{(3)}(x_m) >_\epsilon$ in the $\epsilon \to 0$ limit, $< S^{(i)}(x_n)S^{(i)}(x_m) >_\epsilon$ must present a pole singularity in that limit, i.e. it represents a long range correlation whose quanta are the NG fields $B(x_n)$ and $B^\dagger(x_n)$. These are asymptotic fields corresponding to $S^{(i)}(x_n)$, $i = \pm$ (or $i = 1, 2$) and describe the two collective modes dynamically generated by the symmetry breakdown.

The asymptotic state $|0 >_{\mathcal{M}}$ of the net is thus described in terms of the irreducible set of asymptotic fields $\{\phi(x_n), B(x_n)\}$ and their hermitian conjugates.

It can also be proven (Umezawa et al., 1982; Shah et al., 1974) that when the basic field $\psi(x_n)$ (or $S^{(i)}(x_n)$) undergoes the $SU(2)$ transformations (7) (or (8)) the asymptotic fields $\phi(x_n)$, $B(x_n)$ and $B^\dagger(x_n)$ are transformed as

$$\begin{aligned} \phi(x_n) &\rightarrow \phi'(x_n) = \phi(x_n) \\ B(x_n) &\rightarrow B'(x_n) = B(x_n) + i\theta_1(\frac{M}{2})^{\frac{1}{2}} \\ B^\dagger(x_n) &\rightarrow B'^\dagger(x_n) = B^\dagger(x_n) - i\theta_1(\frac{M}{2})^{\frac{1}{2}} \end{aligned} \tag{12}$$

for $\theta_2 = \theta_3 = 0$,

$$\begin{aligned} \phi(x_n) &\rightarrow \phi^{'}(x_n) = \phi(x_n) \\ B(x_n) &\rightarrow B^{'}(x_n) = B(x_n) - \theta_2(\frac{M}{2})^{\frac{1}{2}} \\ B^\dagger(x_n) &\rightarrow B^{'\dagger}(x_n) = B^\dagger(x_n) - \theta_2(\frac{M}{2})^{\frac{1}{2}} \end{aligned} \tag{13}$$

for $\theta_1 = \theta_3 = 0$, and

$$\begin{aligned} \phi(x_n) &\rightarrow \phi^{'}(x_n) = e^{i\theta_3\lambda_3}\phi(x_n) \\ B(x_n) &\rightarrow B^{'}(x_n) = e^{-i\theta_3}B(x_n) \\ B^\dagger(x_n) &\rightarrow B^{'\dagger}(x_n) = e^{i\theta_3}B^\dagger(x_n) \end{aligned} \tag{14}$$

for $\theta_1 = \theta_2 = 0$. The transformations (12)–(14) belong to the $E(2)$ group which is the group contraction of $SU(2)$ (Inönü & Wigner, 1953; De Concini & Vitiello, 1976). It can be shown (Shah et al., 1974) that the field equations for the asymptotic fields $\phi(x_n)$, $B(x_n)$ and $B^\dagger(x_n)$ are invariant under (12–14).

The $B(x_n)$ and $B^\dagger(x_n)$ field translations in Eqs. (12) and (13) describe coherent condensation of these modes in the ground state of the neural net (Umezawa et al., 1982; Klauder & Sudarshan, 1968; Perelomov, 1986). In the infinite volume limit they cannot be induced by any unitary operator, thus these transformations are not unitarily implementable in that limit. States (and the spaces to which they belong) labeled by different values of $\mathcal{M}$ are therefore unitarily inequivalent states (spaces). For more details see the Appendix.

Note that there are no $B$-modes condensed in the normal vacuum state $|0>$, i.e. before the coming of the information input of code $\mathcal{M}$: $<0|B^\dagger(x_n)B(x_n)|0>=0$ for any $n$, since $B(x_n)|0>=0$.

Equations (12) and (13) show that after the input: $<0|B^{'\dagger}(x_n)B^{'}(x_n)|0>= \theta_1^2 M/2$, with $\theta_2 = \theta_3 = 0$, and similar one for $\theta_1 = \theta_3 = 0$; so that, writing in terms of the transformed state $|0>_\mathcal{M}$, ${}_\mathcal{M}<0|B^\dagger(x_n)B(x_n)|0>_\mathcal{M}= \theta_1^2 M/2$, for $\theta_2 = \theta_3 = 0$, and ${}_\mathcal{M}<0|B^\dagger(x_n)B(x_n)|0>_\mathcal{M}= \theta_2^2 M/2$, for $\theta_1 = \theta_3 = 0$. These relations show that $\mathcal{M}$ provides a measure of the number of B-modes condensed in the state of the net once the information carried by the input has been printed there, which corresponds to dynamical spanning of the net by the coherent correlation waves.

Equation (14) show that the state of the net is invariant under rotations around the 3rd axis, in fact $< 0|B^{'\dagger}(x_n)B^{'}(x_n)|0 >= 0 = _{\mathcal{M}} < 0|B^{\dagger}(x_n)B(x_n)|0 >_{\mathcal{M}}$ for $\theta_1 = \theta_2 = 0$.

The physical meaning of the dynamical rearrangement of the symmetry $SU(2) \rightarrow E(2)$ expressed by Eqs. (12)–(14) is the following. The "flip" $u/d$ (or $on/off$) of the neural unit field $\psi(x_n)$ is induced at each site by the $SU(2)$ generators $S^{(i)}(x_n)$, $i = 1, 2$; when however the neural net is acted upon by an external information input of code $\mathcal{M}$, the neural units get correlated over long range and the $SU(2)$ symmetry is broken. The flipping of the unit is then described *not* as an $SU(2)$ rotation but in terms of boson condensation (translation) of $B$ and $B^{\dagger}$ (Eqs. (12) and (13)). In other words, under the external input the neural net is involved as a whole dynamical system and thus local site flipping and fluctuations sum up into collective modes of the net (the $B$-modes): from the point of view of the "memory state" of the net, *the dynamically interesting objects are not the local flipping or fluctuations but the (long range) dynamical correlations among the neural units.*

## 4 Dissipation and Thermal Properties

I remark that once an information has been stored in the net, the state in which the net was *before* the storage of the information cannot be recovered anymore unless an external operator *resets* the system, which is a quite "dramatic" event for the net, indeed.

The dynamics of information storage is thus an irreversible one (Vitiello, 1995); in other words, the net dynamics is not invariant under inversion of time: time-reversal symmetry is broken by the information storage process. This means that our modeling has to be extended to dissipative dynamics.

The canonical formalism can only describe *closed* systems. The study of dissipative systems requires that not only the system of interest has to be considered, but also the environment (or thermal bath) with which the system exchanges energy, particles, etc.. By extending the dynamics so to include also the environment, the *open* system under study and the environment in which it is embedded form together a global *closed* system.

Since the only requirement to be satisfied at the equilibrium is the in/out-energy balance (for simplicity we may assume no particles are exchanged), the environment needs not to be represented in its details except for what concerns the energy fluxes (Vitiello, 1995; Celeghini et al., 1992). Then, in a standard fashion, the environment may be described as the double or "time-reversed" copy of the given system (Vitiello, 1995; Celeghini et al., 1992), so that damping solutions for the system correspond to growing solutions for the environment, and vice-versa. In formal terms, the doubling of the degrees of freedom of the system is introduced in a way that the total {system-environment} energy is represented by the *difference* between the respective energies (Vitiello, 1995; Celeghini et al., 1992).

Actually, one should start since the beginning with doubling the basic fields $\psi$ and work out the full dynamical problem. However, for simplicity it is enough to consider here the doubling of the asymptotic fields, say $\tilde{B}$, $\tilde{B}^\dagger$ and $\tilde{\phi}$. The tilde-fields (the "mirror" modes) $\tilde{B}$ and $\tilde{B}^\dagger$ commute with $B$ and $B^\dagger$; $\tilde{\phi}$ and $\phi$ are anticommuting fields.

Let me first consider the $B$-modes since their coherent boson condensation describes the information storage in the net. Later I will discuss the $\phi$-fields. In the following, the notation and the formalism developed in Vitiello (1995) and Celeghini et al. (1992) is closely followed.

Let $B_{\mathbf{k}}$ ($\tilde{B}_{\mathbf{k}}$) and $B_{\mathbf{k}}^\dagger$ ($\tilde{B}_{\mathbf{k}}^\dagger$) denote the annihilator and the creation operators of the $B$ ($\tilde{B}$) field (see the Appendix). The set of simultaneous eigenvectors of $\hat{N}_{B_{\mathbf{k}}} \equiv B_{\mathbf{k}}^\dagger B_{\mathbf{k}}$ and $\hat{N}_{\tilde{B}_{\mathbf{k}}} \equiv \tilde{B}_{\mathbf{k}}^\dagger \tilde{B}_{\mathbf{k}}$ is denoted by $\{|\mathcal{N}_{B_{\mathbf{k}}}, \mathcal{N}_{\tilde{B}_{\mathbf{k}}} >\}$, with $\mathcal{N}_{B_{\mathbf{k}}}$ and $\mathcal{N}_{\tilde{B}_{\mathbf{k}}}$ non-negative integers, and $|0>_0 \equiv |\mathcal{N}_{B_{\mathbf{k}}} = 0, \mathcal{N}_{\tilde{B}_{\mathbf{k}}} = 0>$ so that $B|0>_0 = 0 = \tilde{B}_{\mathbf{k}}|0>_0$.

The Hamiltonian ruling the dynamics of an (infinite) collection of damped harmonic oscillators is given by

$$H = H_0 + H_I, \tag{15}$$

$$H_0 = \sum_{\mathbf{k}} \hbar\Omega_{\mathbf{k}} \left( B_{\mathbf{k}}^\dagger B_{\mathbf{k}} - \tilde{B}_{\mathbf{k}}^\dagger \tilde{B}_{\mathbf{k}} \right), \tag{16}$$

$$H_I = i \sum_{\mathbf{k}} \hbar\Gamma_{\mathbf{k}} \left( B_{\mathbf{k}}^\dagger \tilde{B}_{\mathbf{k}}^\dagger - B_{\mathbf{k}} \tilde{B}_{\mathbf{k}} \right), \tag{17}$$

where $\Omega_{\mathbf{k}}$ is the frequency (including also the chemical potential) and $\Gamma_{\mathbf{k}}$ is the coupling constant.

The neural net memory state is defined to be a zero energy eigenstate of $H_0$ (the vacuum or ground state). At a given initial time $t_0 = 0$, it is therefore a condensate of *equal number* of modes $B_{\mathbf{k}}$ and mirror modes $\tilde{B}_{\mathbf{k}}$ for any $\mathbf{k}$.

Since the code $\mathcal{M}$ is given by the condensate modes, $\mathcal{M}$ is actually determined by the set of integers $\{\mathcal{N}_{B_{\mathbf{k}}} = \mathcal{N}_{\tilde{B}_{\mathbf{k}}}, \forall \mathbf{k}, at \;\; t_0 = 0\}$ defining the "initial value" of the condensate at time $t_0 = 0$ (in the following such a set will be simply denoted by $\mathcal{M}$). Infinitely many memory states at $t_0 = 0$, each one corresponding to a different number $\mathcal{N}_{B_{\mathbf{k}}}$ of $B_{\mathbf{k}}$ modes, for all $\mathbf{k}$, thus exist, provided $\mathcal{N}_{B_{\mathbf{k}}} - \mathcal{N}_{\tilde{B}_{\mathbf{k}}} = 0$ for all $\mathbf{k}$.

At finite volume $V$, the neural net memory state $|0>_{\mathcal{M}}$ can be then represented as a generalized $SU(1,1)$ coherent state (Vitiello, 1995; Celeghini et al., 1992):

$$|0>_{\mathcal{M}} = \prod_{\mathbf{k}} \frac{1}{\cosh\theta_{\mathbf{k}}} \exp\left( \tanh(-\theta_{\mathbf{k}}) J_+^{(\mathbf{k})} \right) |0>_0 \;\;, \tag{18}$$

with $J_+^{(\mathbf{k})} \equiv B_{\mathbf{k}}^\dagger \tilde{B}_{\mathbf{k}}^\dagger$. Note that $|0>_{\mathcal{M}}$ is normalized to 1 for all $\mathcal{M}$. The modes $B_{\mathbf{k}}$ and $\tilde{B}_{\mathbf{k}}$ are entangled modes (Vitiello, 1995, 2001, 2004a; Freeman & Vitiello, 2006; Pessa & Vitiello, 2004, 2003; Sabbadini & Vitiello, 2019).

In Eq. (18) the set $\theta \equiv \{\theta_{\mathbf{k}}\}$ is related to the $\mathcal{M}$-set, $\mathcal{M} \equiv \{\mathcal{N}_{B_{\mathbf{k}}} = \mathcal{N}_{\tilde{B}_{\mathbf{k}}}, \forall \mathbf{k}, at \;\; t_0 = 0\}$, by

$$\mathcal{N}_{B_{\mathbf{k}}} = {}_{\mathcal{M}}< 0|B_{\mathbf{k}}^\dagger B_{\mathbf{k}}|0>_{\mathcal{M}} = \sinh^2 \theta_{\mathbf{k}} \,, \tag{19}$$

and the notation $\mathcal{N}_{B_{\mathbf{k}}}(\theta) \equiv \mathcal{N}_{B_{\mathbf{k}}}$ is also used. A condition on the $\theta$-set is that $B$ and $\tilde{B}$ modes satisfy the Bose distribution at time $t_0 = 0$:

$$\mathcal{N}_{B_{\mathbf{k}}}(\theta) = \sinh^2 \theta_{\mathbf{k}} = \frac{1}{e^{\beta E_{\mathbf{k}}} - 1} \,, \tag{20}$$

where $\beta \equiv 1/k_B T$ denotes the inverse temperature at time $t_0 = 0$ and $k_B$ is the Boltzmann constant. Equation (20) shows that the set $\theta = \theta(\beta)$, with $\theta_{\mathbf{k}} = \theta_{\mathbf{k}}(\beta)$, $\forall \mathbf{k}$ .

Equation (20) shows that the neural net memory state $|0>_{\mathcal{M}}$ is a finite temperature state (Vitiello, 1995). Moreover, we will see in the next section that (20) is implied by the minimization of the free energy (and vice-versa).

In other words, the QFT dissipative dynamical formalism naturally leads us to the study of thermodynamic properties for the memory recording process. Moreover, inspection of Eq. (18) leads us to recognize that the net memory state is equivalent to the Thermo Field Dynamics representation $\{|0(\theta(\beta)) >\}$ for QFT at finite temperature (Umezawa et al., 1982; Takahashi & Umezawa, 1975; Umezawa, 1993).

As already observed, in the infinite volume limit $\{|0>_{\mathcal{M}}\}$ and $\{|0>_{\mathcal{M}'}\}$, with $\mathcal{M} \neq \mathcal{M}'$, are Hilbert spaces (representations of the canonical commutation relations (*CCR*)) each other unitarily inequivalent, which means that it does not exist any unitary transformation which maps $\{|0>_{\mathcal{M}}\}$ to $\{|0>_{\mathcal{M}'}\}$.

Also note that $[\, H_0, H_I \,] = 0$, which ensures that the number $(\mathcal{N}_{B_{\mathbf{k}}} - \mathcal{N}_{\tilde{B}_{\mathbf{k}}})$ is a constant of motion for any $\mathbf{k}$. Thus, although $\mathcal{N}_{B_{\mathbf{k}}}$ and $\mathcal{N}_{\tilde{B}_{\mathbf{k}}}$ are allowed to separately change in time, their difference is kept constantly zero during time evolution.

We thus have infinitely many degenerate vacua $|0>_{\mathcal{M}}$, for all $\mathcal{M}$, and correspondingly the space of states of the net at $t_0 = 0$ splits into infinitely many unitarily inequivalent representations of the *CCR*'s ("replicas" of the system). A large number of sequentially recorded information may *coexist* without destructive interference since infinitely many vacua $|0>_{\mathcal{M}}$ are independently accessible. Storage of information of code $\mathcal{M}'$ does not necessarily produce destruction of previously stored information of code $\mathcal{M} \neq \mathcal{M}'$. The existence of the degenerate vacua due to dissipativity thus guaranties a huge memory capacity.

The neural net state may then be represented as the superposition of the full set of memory states $|0>_{\mathcal{M}}$, for all $\mathcal{M}$.

The effect of finite (realistic) size of the net (finite number of neural units) may however destroy or reduce unitary inequivalence and may lead to “dirty” or “noisy” information storage (*interference* or *overlap* among information).

Formally, at finite volume $V$, the time evolution of the memory state $|0>_{\mathcal{M}}$, denoted by $|0(t)>_{\mathcal{M}}$ and specified by the initial value $\mathcal{M}$ of the condensate, is given by

$$|0(t)>_{\mathcal{M}} = \exp\left(-it\frac{H}{\hbar}\right)|0>_{\mathcal{M}} \ . \tag{21}$$

Here the notation is $|0(t)>_{\mathcal{M}} \equiv |0(\theta(\beta), t)>_{\mathcal{M}}$ and its explicit expression is the same as the one in Eq. (18) with $(-\theta_{\mathbf{k}})$ replaced by $(\Gamma_{\mathbf{k}}t - \theta_{\mathbf{k}})$.

The state $|0(t)>_{\mathcal{M}}$ can be shown to decay according to the law

$$\lim_{t\to\infty} {}_{\mathcal{M}}<0(t)|0>_{\mathcal{M}} \propto \lim_{t\to\infty} \exp\left(-t\sum_{\mathbf{k}}\Gamma_{\mathbf{k}}\right) = 0 \ , \tag{22}$$

provided $\sum_{\mathbf{k}}\Gamma_{\mathbf{k}} > 0$. In the infinite volume limit, for $\int d^3\mathbf{k}\,\Gamma_{\mathbf{k}} > 0$ and finite, we have

$${}_{\mathcal{M}}<0(t)|0>_{\mathcal{M}} \xrightarrow[V\to\infty]{} 0 \quad \forall t \ , \tag{23}$$

$${}_{\mathcal{M}}<0(t)|0(t')>_{\mathcal{M}} \xrightarrow[V\to\infty]{} 0 \quad \forall t, t' \quad , \quad t \neq t' \ . \tag{24}$$

This means that in the infinite volume limit, time evolution of $|0>_{\mathcal{M}}$ is rigorously frozen (the states $|0(t)>_{\mathcal{M}}$, $|0(t)>_{\mathcal{M}}$ and the associated Hilbert spaces are each other unitarily inequivalent for different time values $t \neq t'$ in the infinite volume limit); however, a finite life-time may be possible due to effects of the system finiteness (finite number of neural units).

Time evolution of the memory state $|0>_{\mathcal{M}}$ is thus represented as the trajectory of “initial condition” specified by the $\mathcal{M}$-set in the space of the representations $\{|0(t)>_{\mathcal{M}}\}$ of the *CCR*’s. The non-unitary character of time-evolution implied by dissipation is recovered in the unitary inequivalence among representations at different times in the infinite-volume limit. Note also that

$${}_{\mathcal{M}}<0(t)|0>_0 = \exp\left(-\sum_{\mathbf{k}}\ln\cosh\left(\Gamma_{\mathbf{k}}t - \theta_{\mathbf{k}}\right)\right). \tag{25}$$

Thus at time $t = \tau$, with $\tau$ the largest of the values $\tau_{\mathbf{k}} \equiv \theta_{\mathbf{k}}/\Gamma_{\mathbf{k}}$, the memory state $|0>_{\mathcal{M}}$ is reduced (decayed) to the “empty” vacuum $|0>_0$: the information has been *forgotten*.

At each instant $t$ the number of modes $B_{\mathbf{k}}$ is given by

$$\mathcal{N}_{B_{\mathbf{k}}}(\theta, t) \equiv {}_{\mathcal{M}}<0(t)|B_{\mathbf{k}}^{\dagger}B_{\mathbf{k}}|0(t)>_{\mathcal{M}} = \sinh^2\big(\Gamma_{\mathbf{k}}t - \theta_{\mathbf{k}}\big) \tag{26}$$

and a similar expression holds for the modes $\tilde{B}_{\mathbf{k}}$. It is remarkable that in computing the number of the $B_{\mathbf{k}}$ modes, the only contributions come from the mirror modes $\tilde{B}_{\mathbf{k}}$ (the same holds for Eq. (19)).

Equation (26) confirms that the information code is lost after a time $t = \tau$.

## 5 Entropy and Free Energy

In this section, some more details of the thermal properties of the neural net are presented. The state $|0(t) >_{\mathcal{M}}$ can be written as (Vitiello, 1995; Umezawa et al., 1982)

$$|0(t) >_{\mathcal{M}} = \exp\left(-\frac{1}{2}S_B\right)|\,\mathcal{I} > = \exp\left(-\frac{1}{2}S_{\tilde{B}}\right)|\,\mathcal{I} > \;, \tag{27}$$

with $|\,\mathcal{I} > \equiv \exp\left(\sum_{\mathbf{k}} B^{\dagger}_{\mathbf{k}} \tilde{B}^{\dagger}_{\mathbf{k}}\right)|0>_0$ and

$$S_B \equiv -\sum_{\mathbf{k}} \Big\{ B^{\dagger}_{\mathbf{k}} B_{\mathbf{k}} \ln \sinh^2\big(\Gamma_{\mathbf{k}} t - \theta_{\mathbf{k}}\big) - B_{\mathbf{k}} B^{\dagger}_{\mathbf{k}} \ln \cosh^2\big(\Gamma_{\mathbf{k}} t - \theta_{\mathbf{k}}\big) \Big\}. \tag{28}$$

A similar expression is obtained for $S_{\tilde{B}}$ with $\tilde{B}_{\mathbf{k}}$ and $\tilde{B}^{\dagger}_{\mathbf{k}}$ replacing $B_{\mathbf{k}}$ and $B^{\dagger}_{\mathbf{k}}$, respectively. I shall simply write $S$ for either $S_B$ or $S_{\tilde{B}}$. $S$ represent the entropy operator for the dissipative system (Umezawa et al., 1982; Celeghini et al., 1992).

The variation of $|0(t) >_{\mathcal{M}}$ in time, at finite volume $V$, is given by

$$\frac{\partial}{\partial t}|0(t) >_{\mathcal{M}} = -\left(\frac{1}{2}\frac{\partial S}{\partial t}\right)|0(t) >_{\mathcal{M}} . \tag{29}$$

Thus, $i\hbar\,(1/2)(\partial S/\partial t)$ acts as the generator of time-translations: time-evolution is controlled by the entropy variations; the operator $S$ that controls time evolution also defines the system entropy. This indeed reflects the irreversibility of time evolution (breakdown of time-reversal symmetry); namely, the choice of a privileged direction in time evolution (*arrow of time*). The free energy functional is

$$\mathcal{F}_B \equiv {}_{\mathcal{M}}< 0(t)|\Big(H_B - \frac{1}{\beta}S_B\Big)|0(t) >_{\mathcal{M}} \;. \tag{30}$$

$\beta$ is the inverse temperature: $\beta(t) = 1/k_B T(t)$; $H_B$ is the Hamiltonian for the $B$-modes only, $H_B = \sum_{\mathbf{k}} \hbar\Omega_{\mathbf{k}} B^{\dagger}_{\mathbf{k}} B_{\mathbf{k}}$. Let $\Theta_{\mathbf{k}} \equiv \Gamma_{\mathbf{k}} t - \theta_{\mathbf{k}}$ and $E_{\mathbf{k}} \equiv \hbar\Omega_{\mathbf{k}}$. The condition

$$\frac{\partial \mathcal{F}_B}{\partial \Theta_{\mathbf{k}}} = 0\,, \quad \forall\, \mathbf{k}\,, \tag{31}$$

gives $\beta(t)E_{\mathbf{k}} = -\ln\tanh^2(\Theta_{\mathbf{k}})$, i.e

$$\mathcal{N}_{B_{\mathbf{k}}}(\theta, t) = \sinh^2\bigl(\Gamma_{\mathbf{k}}t - \theta_{\mathbf{k}}\bigr) = \frac{1}{\mathrm{e}^{\beta(t)E_{\mathbf{k}}} - 1} \,, \tag{32}$$

which is the Bose distribution for $B_{\mathbf{k}}$ at time $t$.

The entropy $\mathcal{S}(t) = {}_{\mathcal{M}} < 0(t)|S|0(t) >_{\mathcal{M}}$, for $B$ and $\tilde{B}$ modes grows monotonically with $t$ from value 0 at $t = \tau$ to infinity at $t = \infty$ . The difference $(S_B - S_{\tilde{B}})$ however is constant in time, as expected since the role of $\tilde{B}$-modes is in fact the one of closing the (open) system of $B$-modes.

$\mathcal{S}(t)$ is a decreasing function of time in the interval $(t_0 = 0, \tau)$. This ensures that the memory state, is protected from cancellation, although not conserved in time. In this process, of course, the energy exchange with the environment is crucial and finite volume effects are also assumed. In the infinite volume limit, as already noticed, stability is rigorously ensured due to unitary inequivalence between memory states with $\mathcal{M} \neq \mathcal{M}'$.

When changes in temperature are negligible, $\partial\beta/\partial t = -(1/k_B T^2)(\partial T/\partial t) \approx 0$, the change in the energy $E_B \equiv \sum_{\mathbf{k}} E_{\mathbf{k}}\mathcal{N}_{B_{\mathbf{k}}}$ and in the entropy is given by

$$dE_B = \sum_{\mathbf{k}} E_{\mathbf{k}}\dot{\mathcal{N}}_{B_{\mathbf{k}}}dt = \frac{1}{\beta}d\mathcal{S}_B \,, \tag{33}$$

i.e.

$$dE_B - \frac{1}{\beta}d\mathcal{S}_B = 0 \,, \tag{34}$$

which expresses the first principle of thermodynamics for a system coupled with environment at constant temperature and in absence of mechanical work. As usual heat is defined by $dQ = (1/\beta)dS$ and thus Eq. (33) shows that the change in time of condensate turns out into heat dissipation $dQ$.

Let us now consider the $\phi$ and the $\tilde{\phi}$ fields. They behave as anticommuting (fermion) fields and also for them the total Hamiltonian is the difference between the respective ones for $\phi$ and for $\tilde{\phi}$ (Umezawa et al., 1982; Takahashi & Umezawa, 1975; Umezawa, 1993).

Let $a_{i,\mathbf{k}}$, $\tilde{a}_{i,\mathbf{k}}$, and $a^{\dagger}_{i,\mathbf{k}}$, $\tilde{a}^{\dagger}_{i,\mathbf{k}}$, $i = u, d$, be the respective annihilation and creation operators. In a way similar to the case of the boson modes $B$ and $\tilde{B}$, one may construct the set $\{|\mathcal{N}_{a_{i,\mathbf{k}}}, \mathcal{N}_{\tilde{a}_{i,\mathbf{k}}} >\}$ of simultaneous eigenvectors of $\hat{\mathcal{N}}_{a_{i,\mathbf{k}}} \equiv a^{\dagger}_{i,\mathbf{k}}a_{i,\mathbf{k}}$ and $\hat{\mathcal{N}}_{\tilde{a}_{i,\mathbf{k}}} \equiv \tilde{a}^{\dagger}_{i,\mathbf{k}}\tilde{a}_{i,\mathbf{k}}$, with $\mathcal{N}_{a_{i,\mathbf{k}}}$ and $\mathcal{N}_{\tilde{a}_{i,\mathbf{k}}}$ equal to zero or to one.

The state $|0>_0$ is defined as $|0>_0 \equiv |\mathcal{N}_{a_{i,\mathbf{k}}} = 0, \mathcal{N}_{\tilde{a}_{i,\mathbf{k}}} = 0, \mathcal{N}_{B_{\mathbf{k}}} = 0, \mathcal{N}_{\tilde{B}_{\mathbf{k}}} = 0;\ \forall\mathbf{k} >$ such that $B_{\mathbf{k}}|0>_0 = 0 = \tilde{B}_{\mathbf{k}}|0>_0$ and $a_{i,\mathbf{k}}|0>_0 = 0 = \tilde{a}_{i,\mathbf{k}}|0>_0$.

The neural net memory state is then the zero energy eigenstate (the vacuum or ground state) such that at a given initial time $t_0 = 0$, it is a condensate of *equal number* of modes $a_{i,\mathbf{k}}$, $B_{\mathbf{k}}$ and mirror modes $\tilde{a}_{i,\mathbf{k}}$, $\tilde{B}_{\mathbf{k}}$ for any $\mathbf{k}$, respectively.

Infinitely many memory states thus exist at $t_0 = 0$, each one corresponding to a different number $\mathcal{N}_{B_{\mathbf{k}}}$ of $B_{\mathbf{k}}$ modes, for all $\mathbf{k}$, and $\mathcal{N}_{a_{i,\mathbf{k}}}$ of $a_{i,\mathbf{k}}$ modes, for all $\mathbf{k}$ and for all $i$, provided $\mathcal{N}_{B_{\mathbf{k}}} - \mathcal{N}_{\tilde{B}_{\mathbf{k}}} = 0$ for all $\mathbf{k}$ and $\mathcal{N}_{a_{i,\mathbf{k}}} - \mathcal{N}_{\tilde{a}_{i,\mathbf{k}}} = 0$ for all $\mathbf{k}$ and for all $i$.

Let $\mathcal{M} \equiv \{\mathcal{N}_{a_{i,\mathbf{k}}} = \mathcal{N}_{\tilde{a}_{i,\mathbf{k}}}, \mathcal{N}_{B_{\mathbf{k}}} = \mathcal{N}_{\tilde{B}_{\mathbf{k}}}, \forall \mathbf{k}, \forall i, \ at \ \ t_0 = 0\}$ denote the set of integers defining the "initial value" of the condensate. At finite volume $V$, the neural net memory state $|0>_{\mathcal{M}}$ can be then represented as the tensor product of a $SU(2)$ generalized coherent state for $\phi$-modes and a $SU(1,1)$ generalized coherent state for $B$-modes. For sake of shortness I do not report here its explicit form (but see, e.g. (Umezawa et al., 1982; Perelomov, 1986)). For such a state it is possible to repeat the analysis of time evolution done for the $B$-modes and also in such a case the finite temperature state, both for $\phi$ and for $B$ fields, is obtained.

As seen in Sect. 3, the $\phi$ (and the $\tilde{\phi}$) fields do not transform under the asymptotic field transformations Eqs. (12) and (13). This shows that the process of memory storage is taken in care solely by the condensation (translation) of the $B$ (and $\tilde{B}$) fields. The *on/off* or $u/d$ state of the $\phi$ fields contributes however to the asymptotic (observable) state of the neural net. The expectation value of their number operator in the net memory state is given by the free energy stationary condition:

$$\mathcal{N}_{a_{i,\mathbf{k}}}(\theta, t) = {}_{\mathcal{M}} < 0(t)|a^{\dagger}_{i,\mathbf{k}} a_{i,\mathbf{k}}|0(t) >_{\mathcal{M}} = \frac{1}{e^{\beta(t)E_{i,\mathbf{k}}} + 1}, \tag{35}$$

which is of course the Fermi distribution for $a_{i,\mathbf{k}}$ at time $t$. Here $E_{i,\mathbf{k}}$ is understood to include the "threshold" of the chemical potential $\mu$ ($E_{i,\mathbf{k}} \equiv \epsilon_{i,\mathbf{k}} - \mu$).

The change due to thermal effect on the expectation of finding $\mathcal{N}_{a_{i,\mathbf{k}}}(\theta, t)$ modes at time $t$, assuming it is equal to one at zero temperature and energy $\epsilon_{i,\mathbf{k}} < \mu$ is

$$\Delta \mathcal{N}_{a_{i,\mathbf{k}}}(\theta, t) = 1 - \frac{1}{e^{\beta(t)E_{i,\mathbf{k}}} + 1} = \frac{1}{e^{-\beta(t)E_{i,\mathbf{k}}} + 1}, \tag{36}$$

which gives the sigmoid activation function often adopted in neural nets modeling, also showing its relation with the dissipative dynamics.

## 6 Conclusion

In this article, the mathematical formalism underlying the many-body model of neural networks has been reported. The model and related computer simulations confirming the theoretical scheme have been originally discussed in Pessa and Vitiello (1999).

In the model, the *on/off* switching pulse of the units of the net is characterized by an amplitude and by a phase determined by the emission time and is described by a complex doublet field $\psi(x_n)$, treated as a fermion field.

The fully random distribution of the $u$ and $d$ (*on* and *off*) states of the units is described by the $SU(2)$ symmetry. The information input presented to the net induces spontaneous breakdown of such a symmetry with consequent dynamical generation of long range correlations among the units, according to the Goldstone theorem of QFT.

The state of the net at the level of its units (the *microscopic* state of the net) can be determined only up to fluctuations between the states $u$ and $d$ of the single units. The *macroscopic* state (the memory state) of the net is specified by the value of the order parameter $\mathcal{M}$, which describes the long range correlations among the net units (*collective modes* or *coherent condensation* of the Nambu-Goldstone boson quanta associated to the correlation waves). $\mathcal{M}$ is the *code* specifying the information carried by the input originating the whole process.

The evolution in time $t$ of the macroscopic state of the system is controlled by the entropy, which in turn, at each $t$ enters in the minimization of the free energy and accounts for the changes of the degree of correlations of the units. The time-evolution of the macroscopic states of the net is thus an irreversible time-evolution, which can be shown to be described by classical chaotic trajectories in the memory space $\{|0(t)\ >_{\mathcal{M}}\}$ (Pessa & Vitiello, 2004, 2003; Sabbadini & Vitiello, 2019; Vitiello, 2004b). The net is indeed a dissipative system, open to the environment from which it receives information inputs and to which it addresses its responses. Remarkably, 't Hooft has proposed that the formalism describing classical deterministic dissipative systems manifests itself as a quantum formalism ('t Hooft, 1999; Blasone et al., 2001b). This supports the quantum character of the formalism adopted in the previous sections (Pessa & Vitiello, 1999) and in the theoretical computer science (TCS) analysis discussed in Basti et al. (2017).

One remarkable feature of the QFT formalism is the existence of infinitely many unitarily inequivalent representations of the commutation relations. This implies that in the many-body model of the net different memories may be recorded in different (unitarily inequivalent) states. The dissipative dynamics then allows the access to each of these representation without "memory overprinting", i.e. cancellation of an already recorded memory by a supervening new one.

One further aspect of the model is that recalling a memory is obtained by "reading" its mirror modes (the tilde modes in the notation of Sects. 4 and 5). In this sense, the model uses the free energy minimization as a sort of "truth evaluation function", namely the matching with the mirror modes (replication signal) ensures the reading of the stored information.

Another important role played by the free energy minimization is that it leads to the Bose-Einstein distribution function for the correlation wave modes and to the Fermi-Dirac one for the fermion fields of the units. The sigmoid activation function for neural net units is in this way also obtained.

A final remark is that the memory state of the neural net constructed in this paper may be related to coherent squeezed states of quantum optics (Yuen, 1976; Stoler, 1970), the squeezing parameter being related to the memory code $\mathcal{M}$ (Vitiello, 1995, 2001, 2004a; Freeman & Vitiello, 2006).

## Appendix: The Generators of Eqs. (12)–(13)

The generators of the field translations of $B(x_n)$ and $B^\dagger(x_n)$ in Eqs. (12)–(13) are mathematically not well defined (they give diverging quantities) (Shah et al., 1974); however, these translations can be understood as limits of unitary transformations as

$$B(x_n) \rightarrow \lim_{f(x_n)\rightarrow 1} (B(x_n) + f(x_n) \cdot const), \tag{37}$$

with $f(x_n)$ any summable function which satisfies the asymptotic field equation for $B(x_n)$ and $B^\dagger(x_n)$. The role of $f(x_n)$ is to make mathematically well definite the generators of (12) and (13), where $\theta_1$ and $\theta_2$ are replaced by $\theta_1 f(x_n)$ and $\theta_2 f(x_n)$, respectively. Thus the generators are:

$$S_f^{(1)} = (\frac{M}{2})^{\frac{1}{2}} \sum\nolimits_n [B(x_n) f(x_n) + B^\dagger(x_n) f^*(x_n)], \tag{38}$$

$$S_f^{(2)} = -i(\frac{M}{2})^{\frac{1}{2}} \sum\nolimits_n [B(x_n) f(x_n) - B^\dagger(x_n) f^*(x_n)], \tag{39}$$

$$S^{(3)} = \sum\nolimits_n [\phi^\dagger(x_n)\lambda_3\phi(x_n) - B^\dagger(x_n) B(x_n)], \tag{40}$$

and they are time-independent since $f(x_n)$ satisfies the field equation for $B$-modes. They close the $e(2)$ algebra (cf. with (4)):

$$[S_f^{(1)}, S_f^{(2)}] = iM \sum\nolimits_n |f(x_n)|^2 = const \cdot \mathbf{1}$$

$$[S^{(3)}, S_f^{(1)}] = iS_f^{(2)}, \quad [S^{(3)}, S_f^{(2)}] = -iS_f^{(1)} \,. \tag{41}$$

Equation (41) have been derived by using

$$[B(x_n), B^\dagger(x_m)]_{t_{x_n}=t_{x_m}} = \delta^3_{x_n, x_m}, \tag{42}$$

with

$$[B_{\mathbf{k}}, B^\dagger_{\mathbf{q}}] = \delta(\mathbf{k} - \mathbf{q}). \tag{43}$$

$B_{\mathbf{k}}$ and $B^\dagger_{\mathbf{k}}$ are the annihilation and creation operators:

$$B(x_n) = \int \frac{d^3k}{(2\pi)^{\frac{3}{2}}} B_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}_n - i\omega_k t_n}, \tag{44}$$

$$B^{\dagger}(x_n) = \int \frac{d^3k}{(2\pi)^{\frac{3}{2}}} B^{\dagger}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}_n + i\omega_k t_n}. \tag{45}$$

The integration is confined to the domain $-\frac{\pi}{d} < k_i < \frac{\pi}{d}$ , $i = 1, 2, 3$, with $d$ the spacing between sites.

Equations (38)–(40) express the dynamical rearrangement of symmetry $SU(2) \rightarrow E(2)$ (Shah et al., 1974; De Concini & Vitiello, 1976). When expressed in terms of basic fields $\psi(x_n)$ the generators $S^{(i)}$ satisfy the $su(2)$ algebra (6); when the same generators $S^{(i)}$ are expressed in terms of asymptotic fields $\phi$, $B$ and $B^{\dagger}$, as in Eqs. (38)–(40), the satisfy the $e(2)$ algebra (41).

When the limit $f(x_n) \rightarrow 1$ is taken, the state space where the field is translated is unitary inequivalent (Shah et al., 1974) to the one where the field is not translated or is translated by a different constant. This confirms that different, i.e. unitarily inequivalent, spaces of the neural net states correspond to different values of the order parameter $\mathcal{M}$ (different translations constants of the Goldstone modes $B$ and $B^{\dagger}$).

In conclusion, as a consequence of the spontaneous breakdown of the $SU(2)$ symmetry, induced by the external input, the corresponding information is recorded in the neural net state as a coherent condensation of NG gapless modes. They are dynamically generated massless quanta associated to the long range correlations among the net units and are thus collective modes (propagating over long range distances they span the whole system).

## References

Amit, D. J. (1989). *Modeling brain functions*. Cambridge: Cambridge University Press.

Anderson, P. W. (1984). *Basic notions of condensed matter physics*. New York: Addison-Wesley.

Basti, G., Capolupo, A., & Vitiello, G. (2017). Quantum field theory and coalgebraic logic in theoretical computer science. *Progress in Biophysics and Molecular Biology*, **A130**, 39–52.

Blasone, M., Jizba, P., & Vitiello, G. (2011a). *Quantum field theory and its macroscopic manifestations*. London: Imperial College Press.

Blasone, M., Jizba, P., & Vitiello, G. (2001b). Dissipation and quantization. *Physics Letters A*, *287*, 205–210.

Celeghini, E., Rasetti, M., & Vitiello, G. (1992). Quantum dissipation. *Annals of Physics (N.Y.)*, *215*, 156–170.

De Concini, C., & Vitiello, G. (1976). Spontaneous breakdown of symmetry and group contraction. *Nuclear Physics*, *B116*, 141–156.

Freeman, W. J. (2000). *Neurodynamics: An exploration of mesoscopic brain dynamics*. Berlin: Springer.

Freeman, W. J., & Vitiello, G. (2006). Nonlinear brain dynamics as macroscopic manifestation of underlying many-body dynamics. *Physics of Life Reviews*, *3*, 93–117.

Haken, H. (1991). *Synergetic computers and cognition. A top-down approach to neural nets*. Berlin: Springer-Verlag.

Inönü, E., & Wigner, E. P. (1953). On the contraction of groups and their representations. *Proceedings of the National Academy of Sciences of the United States of America*, *39*, 510–524.

Itzykson, C., & Zuber, J. (1980). *Quantum field theory*. New York: McGraw-Hill Inc.

Jibu, M., & Yasue, K. (1995). *Quantum brain dynamics and consciousness*. Amsterdam, The Netherlands: John Benjamins.

Jibu, M., Pribram, K. H., & Yasue, K. (1996). From conscious experience to memory storage and retrieval: the role of quantum brain dynamics and boson condensation of evanescent photons. *International Journal of Modern Physics B*, *B10*, 1735–1754.

Klauder, J. R., & Sudarshan, E. C. (1968). *Fundamentals of quantum optics*. New York: Benjamin.

Mézard, M., Parisi, G., & Virasoro, M. (1993). *Spin glass theory and beyond*. Singapore: World Sci.

Nielsen, M. A., & Chuang, I. L. (2000). *Quantum computation and quantum information*. Cambridge: Cambridge University Press.

Perelomov, A. (1986). *Generalized coherent states and their applications*. Berlin: Springer-Verlag.

Pessa, E., & Vitiello, G. (1999). Quantum dissipation and neural net dynamics. *Bioelectrochem and Bioenergetics, 48*, 339–342.

Pessa, E., & Vitiello, G. (2003). Quantum noise, entanglement and chaos in the quantum field theory of mind/brain states. *Mind and Matter*, *1*, 59–79.

Pessa, E., & Vitiello, G. (2004). Quantum noise induced entanglement and chaos in the dissipative quantum model of brain. *International Journal of Modern Physics*, *B18*, 841–858.

Pribram, K. H. (1971). *Languages of the brain*. New Jersey: Englewood Cliffs.

Pribram, K. H. (1991). *Brain and perception*. New Jersey: Lawrence Erlbaum.

Ricciardi, L. M., & Umezawa, H. (1967). Brain physics and many-body problems. *Kibernetik*, *4*, 44–48. Reprinted in (2004) *Brain and being* (pp. 255–266). In G. G. Globus, K. H. Pribram, & G. Vitiello (Eds.) Amsterdam, Netherland: John Benjamins.

Sabbadini, S. A., & Vitiello, G. (2019). Entanglement and phase-mediated correlations in quantum field theory. Application to brain-mind states. *Applied Sciences*, *9*, 3203. https://doi.org/10.3390/app9153203

Shah, M. N., Umezawa, H., & Vitiello, G. (1974). Relation among spin operators and magnons. *Physical Review B*, *10*, 4724–4726.

Sivakami, S., & Srinivasan, V. (1983). A model for memory. *Journal of Theoretical Biology*, *102*, 287–294.

Stoler, D. (1970). Equivalence classes of minimum uncertainty packets. *Physical Review*, *D1*, 3217–3219.

Stuart, C. I. J., Takahashi, Y., & Umezawa, H. (1978). On the stability and non-local properties of memory. *Journal of Theoretical Biology*, *71*, 605–618.

Stuart, C. I. J., Takahashi, Y., & Umezawa, H. (1979). Mixed system brain dynamics: neural memory as a macroscopic ordered state. *Foundations of Physics*, *9*, 301–327.

Takahashi, Y., & Umezawa, H. (1975). Thermo field dynamics. *Collective Phenomena*, *2*, 55–80. Reprinted in *Int. J. Mod. Phys. B*, *10*, 1755-1805 (1996).

’t Hooft, G. (1999). Quantum gravity as a dissipative deterministic system. *Classical and Quantum Gravity*, *16*, 3263–3279.

Umezawa, H. (1993). *Advanced field theory: micro, macro and thermal concepts*. New York: American Institute of Physics.

Umezawa, H., Matsumoto, H., & Tachiki, M. (1982). *Thermo field dynamics and condensed states*. Amsterdam: North-Holland.

Vitiello, G. (1995). Dissipation and memory capacity in the quantum brain model. *International Journal of Modern Physics B*, *9*, 973–989.

Vitiello, G. (2001). *My double unveiled*. Amsterdam, The Netherlands: John Benjamins.

Vitiello, G. (2004a). The dissipative brain. In G. G. Globus, K. H. Pribram, & G. Vitiello (Eds.) *Brain and being* (pp. 315–334). Amsterdam, Netherland: John Benjamins.

Vitiello, G. (2004b). Classical chaotic trajectories in quantum field theory. *International Journal of Modern Physics*, *B18*, 785–792.

Yuen, H. P. (1976). Two-photon coherent states of the radiation field. *Physical Review*, *A13*, 2226–2243.

# Minds and Robots: An Impassable Border

**Paola Zizzi and Massimo Pregnolato**

**Abstract** We present a distinction between the human mind and a robot, mainly based on the presence or absence of a metalanguage. The human mind possesses both metalanguage and formal language (object language), which is a logic, while the robot possesses only the latter, which is provided as a program. The robot cannot use a metalanguage because the latter, devoid of logical rules, is not Turing-computable, and a computer cannot calculate what is incomputable. Metalanguage, which can be seen as the formal language of meta-thought (the thought that thinks of ordinary thought) allows the human mind to overcome the limits of purely mechanical reasoning. This is why a human mind can never be completely reduced to a Turing machine, and instead always will be a robot. Nevertheless, in the quantum case the hypothesis is made that during the programming phase, the programmers mind can become entangled with the quantum robot.

**Keywords** Metalanguage · Meta-thought · Object language · Robots

## 1 Introduction

*In memory of Eliano Pessa*

We humans who hold metalanguage can program a computer/robot that does not have one. A machine uses only the program it is given (the object language).

The reason a computer cannot have its own metalanguage is because it is not algorithmic (it is not Turing-computable). So what did Turing mean by saying that a computer can "think"? He was probably referring to ordinary thinking, which

---

P. Zizzi (✉)
Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy
e-mail: paola.zizzi@unipv.it

M. Pregnolato
Department of Drug Sciences, University of Pavia, Pavia, Italy
e-mail: massimo.pregnolato@unipv.it

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0_15

humans also have, and which is essentially logical and formal. But humans also have meta-thinking, formalized by metalanguage, while computers have only formal ordinary thinking, the program that has been assigned to them.

Eventually Turing realized that to have "complete" intelligence the machine would have to have a human body, and some senses, to be able to interact with the outside world, and hence the idea of artificial intelligence (AI) was born.

We don't know when and if Turing's dream will come true. Certainly, these human-machine interfaces appear to us extremely complicated and difficult to implement right now, but perhaps in a distant future that will be possible ... who knows.

At the moment, we devote ourselves to a purely theoretical and certainly simpler problem, which however is in the field of AI.

We know that Turing treated his machine, the "bombe", created to decode Enigma, as his own creature, and always tried to protect it.

We therefore ask ourselves the following question: is it simply our desire to humanize machines, as we sometimes do with our pets, or is there something more? If we think about it, our animals have become pets because somehow, living with them we have trained them (programmed) and a strong bond has been created. We think that this link can also be established with the machines we program, but how?

Can our metalanguage affect a machine more profoundly than we believe?

The answer is yes in the quantum case. In fact, a quantum metalanguage is the meta-logical description of QFT, essentially restricted to the moment of interaction (in this case man-quantum computer) and therefore, seen by an external observer, it results as a black box. This means that a bond is created during quantum programming but remains hidden. In any case, hidden does not mean non-existent. We cannot observe the influence of this meta-link once the programming has taken place, but in the meantime man and machine have bonded to each other.

This link can be described as entanglement between the statements of the human metalanguage reduced to quantum logical formulas, and the qubits of the machine.

This mechanism is physically described in the context of QFT in a recent paper (Zizzi, 2020b).

In this way the machine has assimilated some of the humanity of those who have programmed it. For this reason, as already highlighted in (Zizzi, 2020d), we believe that an ethics towards machines is necessary, and perhaps Turing had already guessed it.

The Church-Turing Thesis addresses what kinds of numbers humans, or any machine that uses similar logic, can compute. It is a hypothesis about the nature of computable functions. It states that a function on the natural numbers can be calculated by an effective method if and only if it is computable by a Turing machine.

The Turing Test, in which a user having a conversation through a computer tries to determine whether the correspondent on the other end is a person or a program.

In "Intelligent Machinery" (Turing, 1948) Turing asks "whether it is possible for machinery to show intelligent behaviour," and confronts the challenges of "educating" a machine.

It does not seem to us that Turing ever spoke explicitly of artificial consciousness, but only of artificial intelligence, and not quite in the sense of strong AI. For Turing, a machine could be as intelligent as an organized machine can be (that is, well programmed/trained to execute the program correctly).

This is what you understand by reading his original works.

It looks quite strange to us that Turing did not mention metalanguage of Tarski approach to semantic theory of truth (Tarski, 1944).

The Turing test could be much more efficient if it were based on metalanguage, as Searle also did with his "Chinese chamber" test (Searle, 1980), where Searle provides an argument intended to disprove the position of what he named "strong AI".

In our opinion, a new possible test could be conceived as follows. If you tell a joke, where metalanguage is always used, and you test two people, you are sure that the one laughing is a human and the one that doesn't is a robot. If both don't laugh, it means that the human is stupid (he can't use the metalanguage he is provided with) and the test fails. Since this was a joke then, you should have laughed. But even if you're just laughing now, it's still okay.

In this paper, we conjecture that the impassable border between a human mind and a robot, is just metalanguage. Our belief is based, a part from our personal investigations, see for example (Zizzi, 2008, 2020c, d), mainly on Sambin lectures (Sambin, 2007). On this basis, in (Pessa & Zizzi, 2009) it was also conjectured a possible brain-computer interface as a Quantum Cyborg in which a human mind controls, through a quantum metalanguage, the operations of a quantum computer. The reason why computers cannot use a metalanguage is because it is not algorithmic (not Turing-computable) as it has no logical rules. And a computer cannot calculate what is incomputable.

Roger Penrose (1989) was the first to speculate on the non-computational aspects of the mind, based on Gödel's first incompleteness theorem (Gödel, 1931).

Hence, the non-algorithmic side of the human mind has been explored in the depths of quantum logic by one of us (PZ) (Zizzi, 2011a).

The paper is organized as follows.

In Sect. 2 We give a definition of the mind in terms of logical/metalogical modalities, namely classical/quantum logic for ordinary thinking and classical/quantum metalanguage for meta-thought.

In Sect. 3 We introduce the concepts of metalanguage and object language, their relationships and differences.

In Sect. 4 We show that the axiom of identity belongs to the metalanguage, unlike the law of identity, which belongs to the object language. We therefore argue that a robot will never be able to gain self-awareness.

Furthermore, we show that while in a classical metalanguage the axiom of identity is absolute, in a quantum metalanguage it is probabilistic.

In Sect. 5 We discuss, especially in the quantum case, the non-algorithmic aspects of the human mind, where the boundary is found that for a robot is impassable.

In Sect. 6 We present what we call the “pillars” of the human mind, which distinguish it from a robot, which are: Tarski’s truth predicate, the axiom of identity and the cut rule, all three belonging to the metalanguage.

In Sect. 7 We guess that, in the quantum case, during the programming process, the programmer’s mind and the quantum robot get entangled.

In Sect. 8 We review some recent findings in quantum epigenetics and relate them to a novel approach to the non-invasive brain-computer interface based on quantum metalanguage and a theoretical architecture of quantum cyborgs.

Section 9 is a tribute to our friendship with Eliano Pessa, we describe him as a man and as a scientist mainly in the context of AI.

Section 10 is devoted to the conclusions.

## 2 The Mind

In this Section, we will talk about the mind, or rather, how it is understood by us from a formal point of view. We will investigate what the mind is in this sense, and what its modes and patterns of action are. We will ask ourselves if the mind is real, concrete or abstract, and what is the interpretive physical theory of our formal description.

### *2.1 What Is the Mind?*

> *A totally logical mind.*
> *it’s like an all-blade knife.*
> *It makes the hand that uses it bleed.*
> *(Rabindranath Tagore)*

We define Mind as the “Formal Language of Thought”. It is purely abstract.

Our mind can be in two different modes of language: Logic or Meta-logic.

In Logic mode, the mind generally follows a “classical” logic but sometimes it follows a quantum logic, and in such cases we speak of Quantum Mind.

In both cases of Logic mode the mind is algorithmic (Turing-computable) because a logic has logical rules that can be used by a computer. In particular, in the quantum case, the mind has the same logic as quantum computers.

The Meta-logic mode, which controls the logical mode of thinking, has as its formal language a metalanguage, which is not algorithmic because it has no logical rules.

Therefore a computer, both classical and quantum, cannot have a metalanguage because it cannot compute what is not computable. This is the fine line between the mind and computers and it is impassable.

## 2.2 Is the Mind Real?

> *"We are the dreams*
> *of which the void is made"*

The concept of reality, as well as that of truth, when referred to the mind, are "misleading" if they are considered in an "absolute" sense. We should rather associate them with information, through Wheeler's concept of "it from bit" (Wheeler, 1962) or, in the quantum reformulation, "it from qubit" (Zizzi, 2001).

The mind is not the brain: it could be said that brain is the hardware and the mind is the software, but it would fall into a dangerous mind-body, or spirit-matter dualism.

It is more complicated and subtle than that.

It is true that we can think that the hardware consists of some neuronal processes (classical and quantum) which then translate as logical (classical and quantum) gates of the logic (classical and quantum) of the mind (the software).

But it doesn't stop there, these are only the purely computational aspects.

Thought also has a non-algorithmic aspect. Where does the latter come from?

(a) From the dissipative quantum field theory (DQFT) of the brain (Vitiello, 1995).
(b) A bosonic QFT can be described as a quantum metalanguage (QML) (Zizzi, 2011a, 2020a).

    As a metalanguage has no logical rules and therefore is not Turing-computable, it follows that QFT cannot be completely simulated. In particular, the non-computable sector regards the interaction (Zizzi, 2020b).
(c) In the reduction of QFT to quantum mechanics (QM) (Zizzi, 2020b), one can think that this QML is reflected in the quantum logic of the mind.
(d) "Principle of Reflection" (Sambin et al., 2000):

    - The statements of the meta-language (ML) are reflected in the propositions of Logic, the language-object (OL).
    - The metalinguistic links between ML assertions are reflected in the logical connectives between propositions in the OL.

So in the end, by putting together "(a), (b), (c) and (d)" we have the following scheme in Fig. 1:

An important thing to note in the diagram in Fig. 1, exactly in the red arrow, is that what assigns a "status" of (quantum) metalanguage to QFT is precisely the set of non equivalent vacua (Zizzi, 2020b) for the existence in QFT of unitarily inequivalent representations of the canonical commutation relations (CCR).

## 2.3 The Three Modalities of the Mind: A Deeper Insight

Let's make the formal distinction between ordinary thinking and meta-thinking:

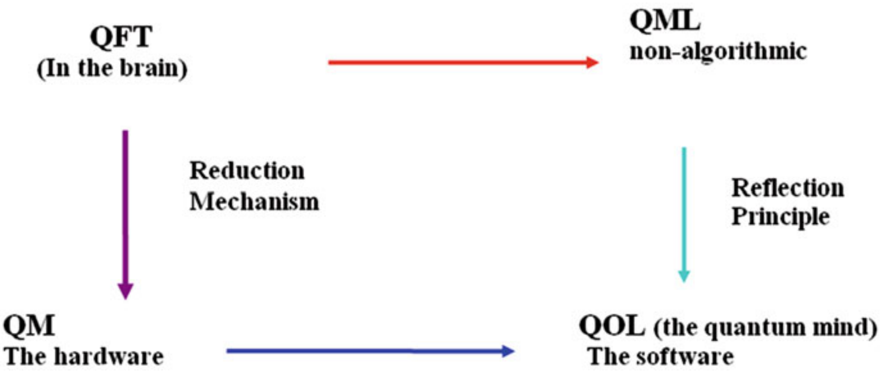


**Fig. 1** *QFT* quantum field theory, *QML* quantum metalanguage, *QM* quantum mechanics, *QOL* quantum object language; On the LHS: the physical theories. On the RHS: the formal languages of the physical theories. The horizontal arrows associate the physical theories to their respective languages. The vertical arrow on the LHS is the reduction mechanism from QFT to QM. The vertical arrow on the RHS is the reflection principle from QML to QOL

*Ordinary thinking:*

1. conscious—classical calculus, classical formal language. We call it "Mind".
2. unconscious-quantum computation, quantum formal language. We call it "Quantum Mind".

*Meta-thinking:*

3. Metalanguage (classical and quantum), non-algorithmic.

We have then three patterns or modalities (Zizzi & Pregnolato, 2012a, 2020):

(a) The quantum modality
(b) The classic modality
(c) The non-algorithmic modality.

Let us start with the quantum modality.

*Ordinary unconscious thinking*: driven by mental processes that are extremely fast, much more so than those involving conscious thinking. This already suggests that the above processes are quantum-computational (a quantum computer is exponentially faster than its classical counterpart). Sudden decision or understanding, creativity, imagination and discoveries, born from an unconscious state of mind, are only the results of a quantum mental process, the intermediate steps of which, however, remain unknowable.

(a) In quantum modality: the result of a quantum computation with a given probability can be obtained, but the intermediate steps are not available. Thus, these two characteristics seem to indicate that the unconscious mind is indeed quantum-computational: the Quantum Mind.

(b) Now, let us consider the classic modality: the unconscious mind calculates in quantum mode and “prepares”, at maximum speed, what we then recognize as conscious thought. Conscious thinking derives from a choice (a measure) made on the quantum computational state, and then uses a classical modality. We don’t have much time to process the outputs of the unconscious mind (half a second), therefore, our conscious thought looks more like a succession of flashes of consciousness rather than a continuous flow. We use partial information obtained from quantum measurements. But in fact, we don’t calculate anything new. Humans calculate quantum, and they don’t have time to realize it.

(c) Finally, we illustrate the non-algorithmic modality. Meta-thinking is the process of thinking about our own thinking. It has no method of calculation, neither classical nor quantum. Quantum meta-thinking, which thinks unconscious quantum thinking, can be seen as the roots of the unconscious mind (the roots of the Quantum Mind). It is the aspect of thought most closely related to matter (physical processes in the brain). The latter should be described by DQFT. Quantum meta-thinking coordinates intuition, intentions and (quantum) control. Meta-thought processes could be interpreted as aiming to maintain a kind of coherence of ordinary thinking (coherent states in DQFT).

## 3 Object Language and Metalanguage: So Closely Related and Yet So Different

The philosophical approach to this chapter, and the reproduction of Figs. 3 and 4 were borrowed from Sambin’s lectures (2007) where you learn logic by teaching it to a robot.

A metalanguage is a language that speaks of another language, called “object language”. When the object language is a formal language such as a logic or a computer program, we say that the corresponding metalanguage is formal.

The distinction between metalanguage and object language is fundamental not only in logic, but also in everyday life. We are constantly at play between the two levels, and we should realize this in order to better understand our own way of thinking.

To get to the metalanguage, which is the most abstract level of reference of thought, we have to go through two lower levels:

In the first place, recognize the expressions (logical formulas in the case of a formal system) that is the most concrete and basic level, which is the one that robots are also equipped with.

Second, give meaning to those expressions and make them propositions (on a more abstract level). This is interpretation: an assignment of meanings to the symbols and words of a language. These two levels are both in the object language, the first is peculiar to machines, which deal only with expressions and formulas, the

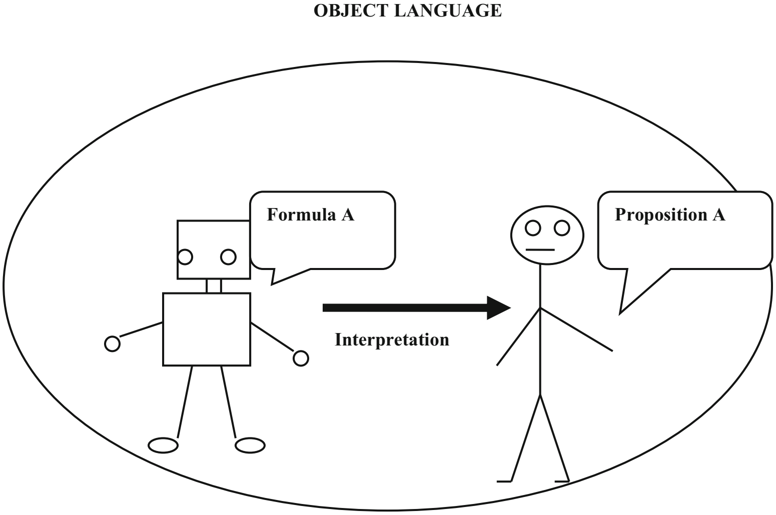


**Fig. 2** Two levels in the object language: the robot “recognizes” the formula (expression) A and, through an interpretation, the man makes it the proposition A

second is the interpretation of these expressions as logical propositions by man. See Fig. 2.

Finally, by declaring (asserting) propositions, we make them assertions, and we enter the even more abstract world of metalanguage.

While we humans have both levels of object language and metalanguage available, a robot has only that of object language at its disposal and is stuck there. See Fig. 3.

The basic elements of a metalanguage are the assertions (asserted propositions) and the metalinguistic links between assertions, which are the metalinguistic “and” denoted by and, and the “yields” (or “entails”) denoted by. In the formalism of sequent calculus (Gentzen, 1969) an assertion Aass. will be indicated with a sequent having the antecedent empty .

Other elements of the metalanguage, always in the framework of sequent calculus, are the axiom of identity and the cut rule. Moreover Tarski truth predicate also stands, together with Tarski Convention T (Tarski, 1944), in the metalanguage. The axiom of identity, the cut rule (Gentzen, 1969), and Tarski convention T will be discussed in the next sections.

To conclude this section, it might be worth discussing compound assertions.

Given two propositions A and B in the object language, they correspond to the assertions A ass. and B ass. in the metalanguage respectively. If we say” A ass” the robot understands “A”, if we say “B ass” the robot understand “B”. But if we say “A ass and B ass” what does the robot understand? We should give him a logical connective & (the logical conjunction, most often denoted by) such that applied to the two propositions A&B produces a new proposition A&B such that:

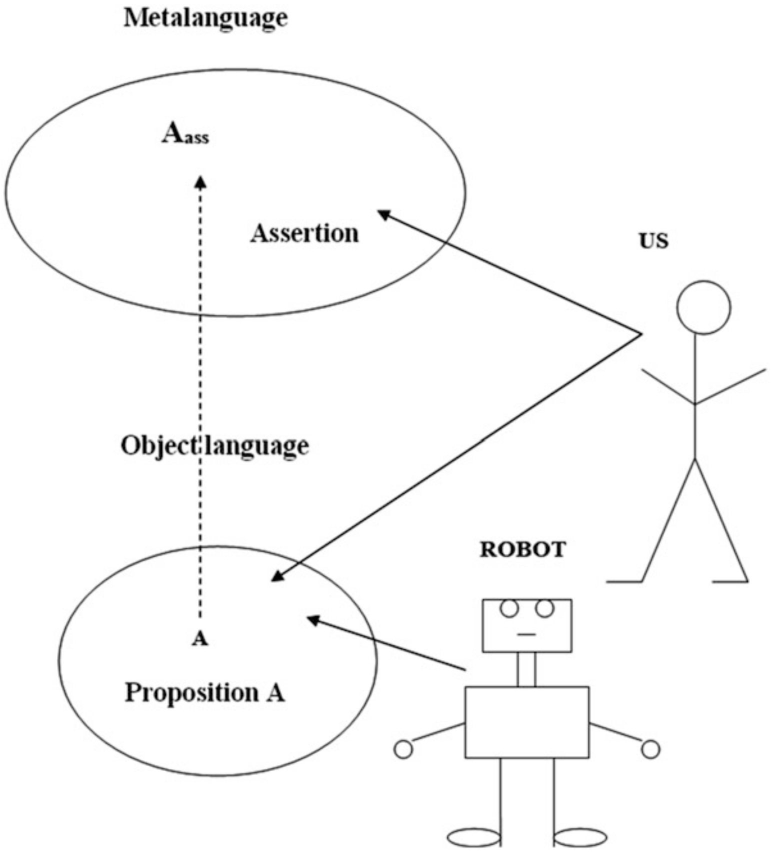


**Fig. 3** Assertions stand in the metalanguage. They are asserted propositions. Propositions stand in the object language. Human beings (US) can reach both levels, robots only that of object language. The subscript "ass" in $A_{ass}$ stands for "A asserted"

A&B ass is equivalent to A ass and B ass.

The above relation produces the "definitional equation" (Sambin et al., 2000) for the logical connective &:

$$|- A \mathbin{\&} B \ \underline{\text{iff}}\ |- A \quad \underline{\text{and}}\ |- B$$

where iff stands for "if and only if".

There exists a definitional equation for every logical connective. Note that what happens is the reflection of the metalinguistic links between assertions into the logical connectives between propositions.

This is called the "reflection principle" (Sambin et al., 2000).

In summary:

A metalanguage (ML) is a language which talks about another language, called object language (OL).

A formal ML consists of assertions, and meta-linguistic links among them. It consists of:

1. Atomic assertions: $|{-}A$ ($A$ declared, or asserted), where $A$ is a proposition of the OL.
2. Meta-linguistic links: $|{-}$ ("yelds", or "entails"), $\underline{\text{and}}$ (metalinguistic "and").
3. Compound assertions. Example: $|{-}A$ $\underline{\text{and}}$ $|{-}B$.

Let us consider the introduction of the logical connective & in Basic logic (Sambin et al., 2000).

In the OL, let $A$, $B$ be propositions.

In the ML, I read: A decl., B decl, that is: $|{-}A\cdots, \cdots|{-}B$ respectively (where "decl." is the abbreviation of "declared", which also can mean "asserted"). Let us introduce a new proposition A&B in the OL. In the ML, we will read: A&B decl., that is: $|{-}A\ \&\ B$. The question is: From *A &B decl.*, can we understand *A decl.* $\underline{\text{and}}$ *B decl.*? More formally, from $|{-}A\ \&\ B$ can we understand $|{-}A$ $\underline{\text{and}}$ $|{-}B$? To be able to understand *A decl.* $\underline{\text{and}}$ *B* decl. From *A&B decl*, we should solve the definitional equation of the connective *&* in Basic logic. See Fig. 4.

## 4 The Disintegrated Self

> *"You are me*
> *And I am you*
> *One is one*
> *And one are two".*
> *I am you. Milonga triste.*

The classical laws of thought are:

Law of identity: $A \rightarrow A$ (states that an object is equal to itself).

Law of the excluded third: $(A \vee \neg A) = 1$ (A or not A is true).

Law of non-contradiction: $(A \wedge \neg A) = 0$ (A and not A is false).

It should be emphasized that the law of identity belongs to logic (the object language). Instead, the axiom of identity:

$$A\,|{-}A$$

belongs to the metalanguage, and it is its reduction (Zizzi, 2020d) to object language which gives rise to the law of identity. We will limit ourselves to the study of the axiom of identity and its psychological interpretation as self-awareness. The derivation of the law of identity in the object language from the identity axiom in the metalanguage was demonstrated in (Zizzi, 2020d).

Here we give only a qualitative explanation in Fig. 5.

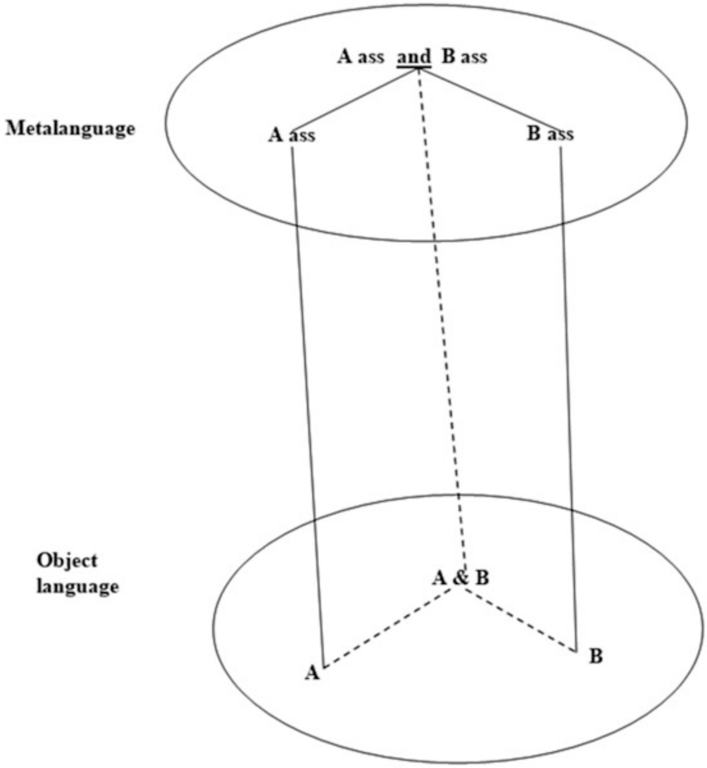


**Fig. 4** The two assertions A $_{\text{ass.}}$ and B $_{\text{ass.}}$ of the metalanguage correspond to the two propositions A and B of the object language. The metalinguistic link and between the two assertions corresponds to the logical connective & between the two propositions

As we will see, in the case of a quantum metalanguage, the axiom of identity is no longer absolute. In the corresponding quantum logic, it follows that an object is only partially equal to itself, the law of non-contradiction is violated, and by duality, also the law of the third excluded is violated.

The classic axiom of identity, which reduces to the classical law of identity in the object language, divides the Universe (U) into two parts: the Self and the Other (Zizzi, 2018). See Fig. 6.

It is a dichotomy: a division of the whole into two parts which are:

Mutually exhaustive $S \cup O = U$ (third party excluded).

Mutually exclusive $S \cap O = \oslash$ (non-contradiction).

The Other is the complement of the Self in U.

In quantum metalanguage, the (classical) axiom of identity is replaced by the quantum one (Zizzi, 2010):

$$A \,\Big|\!-^{|\lambda|^2} A, \quad \alpha \in C$$

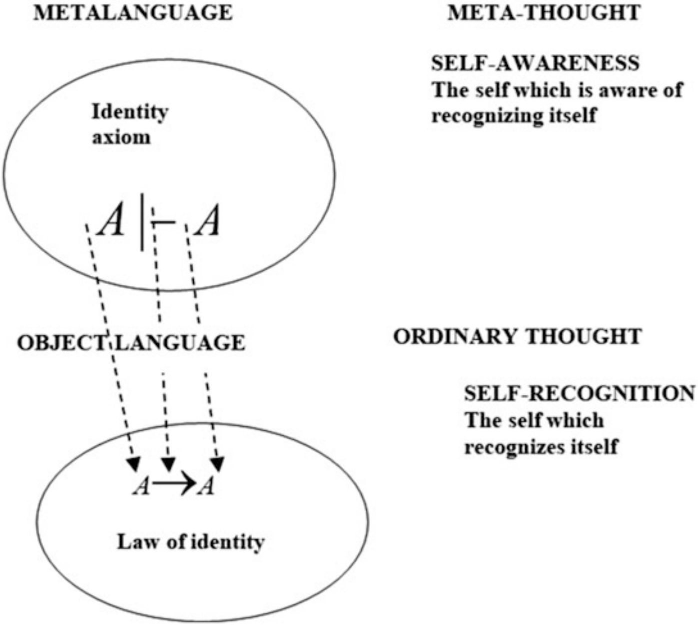


**Fig. 5** The identity axiom stands in the metalanguage, at the level of meta-thought. It represents self-awareness. The law of identity stands in the object language, at the level of ordinary thought

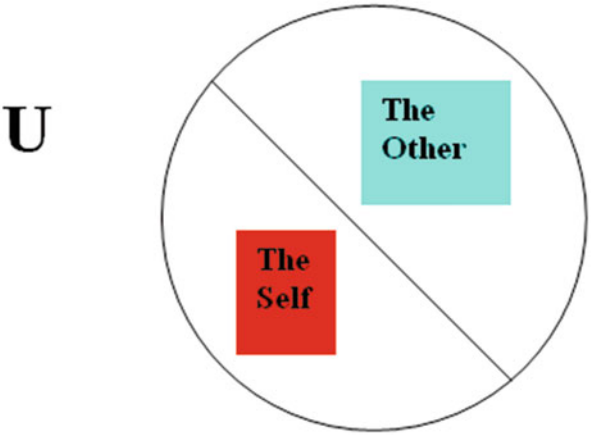


**Fig. 6** Partition of Universe (U) in two parts: The Self and the Other

with partial truth value $v_p = |\lambda|^2 \in [0, 1]$, where $\alpha$ is the degree of the quantum assertion $|-^{\lambda} A$, and corresponds, in quantum mechanics, to a probability amplitude. Therefore the partial truth value $v_p$ corresponds to a probability $p$. This means that a (quantum) object is probabilistically equal to itself.

If that object refers to the Self, the conclusion is that the Quantum Self is "disintegrated" (Zizzi, 2018).

The quantum mental state of a disintegrated self (DS) can be assimilated to a qubit state $|\Psi\rangle_{DS}$ in which the Self is identified with the bit $|1\rangle$ and the Other is identified with the bit $|0\rangle$:

$$|\Psi\rangle_{DS} = \lambda_0 |0\rangle + \lambda_1 |1\rangle, \quad \lambda_0, \lambda_1 \in C, \quad |\lambda_0|^2 + |\lambda_1|^2 = 1.$$

This situation should occur in states of unconsciousness, dreams and schizophrenia (Zizzi & Pregnolato, 2012b).

## 5 The Non-Algorithmic Side of the Mind

*"There are thoughts*
*that are not such"*

Penrose's original conjecture (Penrose, 1989) on the existence of non-algorithmic aspects of the mind was primarily concerned with consciousness. However, (Zizzi & Pregnolato, 2012a) conscious and rational human thought consists of a very rapid sequence of decoherence processes from the quantum to the classical computational mode.

More specifically, in Penrose-Hameroff's Orch-Or theory (Hameroff & Penrose, 1996), overlapping tubulins/qubits decohere and alternate with classical bits at a high rate. According to this theory, it seems that consciousness is made up of "flashes" of classical computation.

The statements of the QML are physically interpreted (Zizzi, 2020b) as quantum fields, in the context of the dissipative quantum field theory (DQFT) of the brain (Vitiello, 1995).

The atomic propositions of the quantum object language (QOL) (Zizzi, 2010) are affirmed, in the quantum metalanguage (QML) with a degree of assertion, which is a complex number.

QML is the language of meta-thought. The very importance of meta-thought, which deals with intuition, intention and control, lies in the fact that it distinguishes man from machines. Indeed, the language of meta-thought, which is non-algorithmic, being described by a metalanguage, cannot be acquired independently by a machine, which is endowed only with an object language.

As is well known, in 1950 A. M. Turing (Turing, 1950) adopted a purely behavioural criterion (instantiated through his famous test) to establish whether a machine could be considered intelligent.

Within this approach, a machine was recognized as having a mind when its behaviour was indistinguishable from that of a human being performing mental operations.

In the 1980s the philosophical considerations already made by Searle (Searle, 1980) and others began to raise serious doubts about the validity of this definition of the mind.

Proper reasoning logic should take into account that humans have basic logical rules and, in general, structural rules are ignored. This requires sub-structural logic, which can be seen as the general platform for any other logic. All these requirements were met in BL (Sambin et al., 2000) in the classical case (or classical way).

A quantum version of BL, called Lq, was introduced in (Zizzi, 2010).

In Lq two new logical connectives have been introduced, the logical connectives "quantum superposition" (the quantum version of the classical conjunction), and "entanglement".

Finally, the probabilistic character of any quantum theory is also present in Lq, since the partial truth values, whose interval is the real interval [0,1], are interpreted as probabilities.

This takes into account the fuzzy and probabilistic character of some non-formalized aspects of thinking. In this context, we will try to clarify the Penrose conjecture (Penrose, 1989) on the non-computational aspects of the mind in relation to Gödel's first incompleteness theorem (Gödel, 1931).

Penrose states that a mathematician can recognize the truth of a Gödel proposition G, although the latter cannot be proved within the axiomatic system, since he is able to recognize an indecipherable truth due to the non-algorithmic aspect of the Mind.

In our opinion, the fact that the mathematician can assert the truth of G is that he is using the non-computable mode of mind described by the metalanguage, where the statements are found and where Tarski introduced the truth predicate (Tarski, 1944).

Furthermore, the fuzzy (Zadeh, 1996) -probabilistic characteristics of the QML lead to modify Tarski Convention T as Convention PT (Zizzi, 2011b), where P stands for "Probably".

There are close relationships between metalanguage assertions, the truth values of propositions in the object-language and Tarski's truth predicate, the latter being formulated in the metalanguage.

However, when the certainty in the statement is not total, the truth values of the propositions are also partial and Tarski's truth predicate must be modified.

With Tarski's convention T, each sentence p of the OL object language must satisfy:

$$(\mathrm{T}) : \text{``}\mathrm{p}\text{''} \text{ is true iff p}$$

where "p" stands for the name of the proposition p, which is the ML metalanguage translation of the corresponding OL proposition, and "iff" stands for "if and only if".

For any "probably p" (P (p)) proposition, we can reformulate Tarski's Convention (T) as a convention (TP) as follows.

$$(\mathrm{TP}):\ \text{“}\mathrm{p}\text{''}\ \text{is probably true iff P}\ (\mathrm{p})$$

The expression “is probably true” means that the truth of a proposition is stated with uncertainty, not with complete certainty. The predicate of truth has been modified by probability.

In the formalism of the calculation of the sequents, the TP convention reads:

$$|\!-^{\lambda}\ {}^{,}\mathrm{p}^{,}\quad \underline{\text{iff}}\quad \mathrm{P}\,(\mathrm{p})$$

which means that the proposition ‘*p*’ is asserted with a degree of assertion λ if and only if “probably *p*”, with probability:

$$|\lambda|^2 \in [0, 1]$$

and the partial truth value of $P(p)$ is just the probability of $p$, that is:

$$v\,(P(p)) = p(p) = |\lambda|^2$$

Practical example: (T) the proposition “the snow is white” is true if and only if the snow is white.

Practical example: (PT) the proposition “the snow is white” is probably true if and only if the snow is probably white (it can have shades).

## 6 The Three Pillars of the Human Mind

Tarski’s truth predicate (both classical and quantum) and the axiom of identity (both classical and quantum) are both formulated in the meta-language, which is not algorithmic. Also, the cut rule, which is a particular rule of sequent calculus, is in fact a meta-rule, that is, a rule that can be formulated only in the metalanguage.

Therefore, these are the three “pillars” of the human mind, which distinguish it from a computer (both classical and quantum).

So a computer/robot, not having the axiom of identity available, will not have self-awareness (Zizzi, 2020d).

Moreover, not having the predicate (T) or (PT) available, he will not be able to be aware of the truth (or falsity) of the external world, therefore of reality itself.

The cut rule is a rule in the sequent calculus-style, which is a generalization of the “modus ponens”: “*P* implies *Q* and *P* is true, therefore *Q* must be true.”

The cut rule is neither an inference rule nor a structural rule, but a meta-rule in the sequent calculus. It reads:

$$\frac{\Gamma\, |\!-A \qquad A\, |\!-B}{\Gamma\, |\!-B}$$

If a formula A appears as a conclusion in one proof $\Gamma|-A$, and as a hypothesis in another $A|-B$, then we can deduce another proof $\Gamma|-B$ in which formula A does not appear.

With the use of the cut rule we humans can "lighten" the premises, and not be forced to use redundant information that can instead be ignored. This corresponds to a "measure of utility" for a "convenient choice" that we make almost unconsciously.

We cannot give the rule of the cut to a robot because it would not know what to do with it. It would not be able to identify and eliminate its own redundant information to achieve a certain result.

The quantum version of the cut rule (Zizzi, 2010) is interpreted as a projective quantum measurement. In this case, the inability of the quantum robot / computer to use the quantum cut rule means that it cannot make a quantum measurement. The fact that a quantum robot cannot perform measurements, had already been pointed out by Benioff (1998).

Finally, they cannot control their quantum object language (the program) that was provided to them by the programmer. One could have hoped that a quantum robot could do it, however .. it did not happen. In 2008 "I, quantum robot" (Zizzi, 2008) was born, but it did not have metalanguage as well.

It might be possible that a quantum metalanguage QML$'$ would be generated as a quantum emergent phenomenon from the quantum object-language QOL (which was induced by the quantum metalanguage QML). However, in this case, the QOL is not anymore active (the quantum machine QM is not anymore a quantum computer) because of Goedel incompleteness theorem, which forbids a formal system (powerful enough to describe arithmetics) to speak about itself. It can happen, nevertheless, that the emergent quantum metalanguage QML$'$ acts as a quantum control on another quantum machine QM$'$, triggering quantum computation (a new quantum object-language QOL$'$). As it should be QOL$' =$ QOL, because all quantum computers share the same quantum logic, one might refer to QML$'$ as it was QML, what is false. Obviously, QML$'$ must be a copy of QML, in order to reflect into QOL$'$ which is identical to QOL, but while the copy QML$'$ is an emergent phenomenon from QOL, the original QML is due to dissipative quantum brain processes. It is possible then that a long sequence of identical quantum metalanguages QML$'$, QML$''$.....is generated and a generation of conscious quantum robots Q$'$, QR$''$, QR$'''$.... QR$^{n}$ come into existence, but when Q$'$ starts its life, QR dies (because as said before, the object-language QOL is deactivated) and so on. At the end, a unique quantum robot QR$^{n}$ survives, which is controlled by an identical copy of the original quantum metalanguage QML derived from high-level thought processes.

Notice that the appearance of a copy of the original metalanguage requires the destruction of the original support of the corresponding object-language QOL, namely of the quantum robot QR. Roughly speaking, a quantum metalanguage (a quantum state of intentional thought) can be copied only if the corresponding quantum computation which was triggered by it is deactivated.

This principle is in agreement with the theorem of no-self replication of quantum machines proved by Pati and Braunstein (2008).

We argue then that self-replication of the support is one of the requirements for being a (quantum) mind.

Physically, this principle can be understood as follows. The QML is made of assertions which are physically interpreted as non-hermitian operators of a dissipative QFT (with an infinite number of degrees of freedom) describing brain processes in the brain. On the other side, the QOL is the quantum logic of the mind, and the corresponding physical theory is QM, with a finite number of degrees of freedom.

The quantum mind has then at its disposal both a non-computational mode (the QML) described in QFT and a quantum-computational mode (the QOL) described in QM. A quantum computer (QC) has only a QOL, and its physical theory is QM. Therefore, a QC cannot reach a QML (a non-algorithmic mode of thought) because it is impossible to go from the finite number of degrees of freedom of QM to the infinite number of those of QFT. That is, a quantum computer will never be able to reach a non-algorithmic mode of thought.

This is the difference between a quantum mind and a quantum computer.

The reflection principle, which transforms a metalanguage into an object language, when applied to a QML of non-algorithmic thought and to a QOL of a quantum-computational mind, needs a physical interpretation. The problem is that QML is described, physically, by a (dissipative) QFT, with an infinite number of degrees of freedom, while the QOL is described by quantum mechanics (with a finite number of degrees of freedom) more precisely, by quantum computing, with quantized information (qubits). This reduction mechanism was found in (Zizzi, 2020b).

## 7 Entanglement Between the Programmer's Mind and the Quantum Robot

We think that entanglement (a quantum correlation) is established between the programmer's mind and the quantum robot during the actual programming phase, as we will illustrate shortly. However, we must immediately clarify that this relationship is very short (it lasts a few milliseconds between the passage of the unconscious state to the conscious one of the programmer) and secret, in the sense that an external observer will never be aware of it.

In (Zizzi, 2020b) we looked for a reduction mechanism from (bosonic) QFT to QM that could reveal QFT's Hidden Quantum Information (HQI). We found that HQI was there and was organized in a quantum network of maximally entangled multipartite states. That was the quantum computational "skeleton" of the original QFT. Since such a "skeleton" is itself a quantum network, it seems that it is right to enter it in a one-to-one correspondence with an external QC to simulate the original QFT. In the reduction mechanism, the degrees of freedom of the quantum fields reduce to a finite number of quantum mechanical states, which are maximally entangled multipartite qubit states.

We think that QFT is meta-logically described (Zizzi, 2020a) by a “quantum metalanguage” (QML) (Zizzi, 2010).

If in particular we consider the Dissipative Quantum Field Theory (DQFT) of the brain (Umezawa, 1993; Umezawa & Vitello, 1985; Vitiello, 1974), then it can be interpreted as the programmer’s quantum metalanguage (PQML) and the reduction mechanism can be viewed as the reflection principle (Sambin et al., 2000) that sends the assertions of the PQML to the propositions (logical formulas) of the program, which is a quantum computational logic Lq (Zizzi, 2010).

It is possible to give a general interpretation of the quantum metalanguage in terms of dynamical processes described by a DQFT. It is to be remarked that in this one introduces a relationship between two entities (the quantum metalanguage and the QFT) each one of which allows an infinite set of different possible representations associated to every symbolic description.

This follows from the fact that QFT deals with infinitely many unitarily inequivalent representations of the canonical commutation relations (CCR). Thus our general interpretation cannot be identified with the usual correspondence rules characterizing the interpretations of logical theories.

The logical quantum network of the program, made of maximal entangled logical qubits, is in a one-to-one correspondence with the physical qubits of the QC. See Fig. 7.

More in detail, in (Zizzi, 2020b) we claimed that a quantum computer can simulate the hidden quantum network (HQN) of the quantum system under study. More precisely, a quantum computer can be programmed to be in a one-to-one correspondence with the HQN.

In the reduction process QFT would appear then as the semantics of the quantum logic underlying the quantum information hidden in it. The reduction process would then play the role of a definitional equation (Sambin et al., 2000), which allows the switch from a metalanguage to an object language (the logic). In particular, the quantum version (Zizzi, 2010) of the definitional equation allows to pass from a QML to the quantum logic of quantum information Lq.

Hence, the metalinguistic links between assertions, which are interpretable as interactions of quantum fields, are sent to logical connectives between propositions, which correspond to quantum correlations such as quantum superposition and entanglement.

In this sense, we say that during the programming process, the programmer’s mind get entangled with the quantum computer/robot. However, this entanglement process cannot be tested by an external observer, to whom the interaction appears as a black box. In fact, this process is revealed only to the internal observer (the programmer’s mind in the quantum logic modality) as the one-to-one correspondence requires the identification of the state space of the QC with the background space (Zizzi, 2005), the latter being a (non-commutative) quantum space.

The programmer anyway is not able to describe in terms of a classical logic his entanglement with the quantum robot, because the experience he had was during

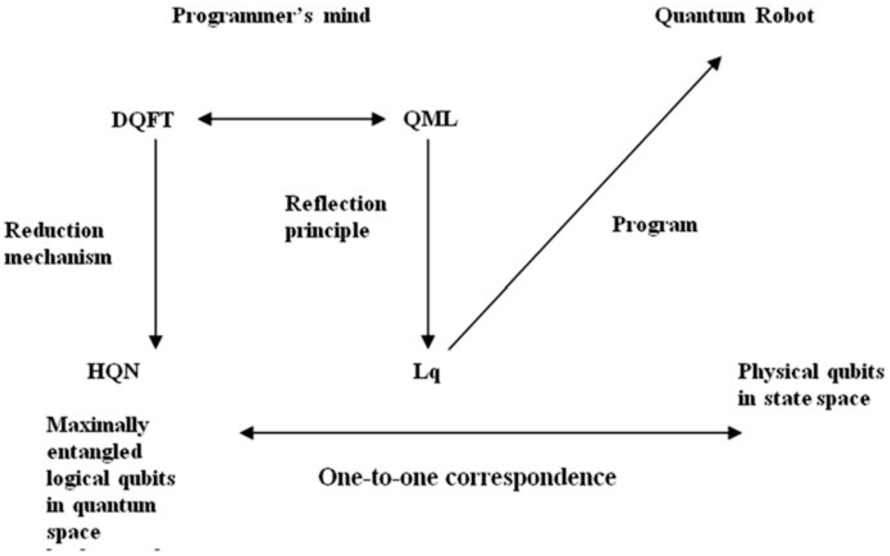


**Fig. 7** On the LHS, the programmer, and the reduction mechanism from DQFT to HQN. In the middle, the corresponding reduction from QML to Lq. On the RHS, the quantum robot (QR). The oblique arrow denotes Lq as the program given to the QR. The horizontal arrow at the bottom denotes the one-to-one correspondence between the background quantum space, and the state space of the QR

a state of unconsciousness, described by a quantum logic Lq (we remind that Lq is the logic of quantum computing, quantum mind or unconscious mind and schizophrenia) (Zizzi & Pregnolato, 2012b).

So this effect will most likely never be used for practical purposes.

However we think that the quantum cyborg is already there, in the programming phase of a quantum robot. If by extrapolation we think of repeated, very fast reprogramming phases, we may 1 day realize that we no longer distinguish the programmer from the quantum robot. This, at least in part, brings back to Turing's idea (Turing, 1948) of "educating" a machine.

## 8 Quantum Cyborgs, Biophotons and Mental Diseases

Quantum robots, originally discussed by Benioff, have no awareness of their environment and do not make decisions or take measurements. We can therefore ask ourselves whether in the future quantum robots will be able to be aware of the environment and perform experiments. This means that they can also become self-aware and have "free will". In the context of a dissipative Quantum Field Theory of brain functioning it is possible to introduce generalized coherent states associated, in the context of logic, with the assertions of a quantum metalanguage.

The latter controls the quantum-mechanical computation corresponding to the standard mental operation.

It thus becomes possible to conceive a Quantum Cyborg in which a human mind controls, through a quantum metalanguage, the operation of an artificial quantum computer.

Classic brain-computer interfaces (cBCI) are systems that acquire and analyze brain signals (typically of an electromagnetic nature) to create real-time broadband communication channels between the human brain and a computer (Pessa & Zizzi, 2009).

A quantum robot (QR) can be defined as a mobile system that has a quantum computer on board and all necessary auxiliary systems. A QR moves and interacts with the environment of a quantum system. In their article Pessa and Zizzi discussed the possibility of implementing a QR with a new and powerful BCI that allows quantum computer—quantum computer communication. The whole system consisting of a human subject and an artificial quantum computer (controlled by the subject's quantum metalanguage) is a new type of cyborg, called Quantum Cyborg (QC). A human subject, through quantum metalanguage, could guide a QC, through a new BCI (much more powerful than the existing ones), transforming it into a more effective direct action of the mind on matter.

Biophotons are mainly produced by molecular species electronically excited in numerous oxidative metabolic processes in cells. They can play a role in cell-to-cell communication and have been observed in different parts of the body, including the brain. Photons in the brain could be ideal candidates for information transfer. They travel tens of millions of times faster than a typical electrical neural signal and are not prone to thermal noise at body temperature due to their relatively high energies (Kumar et al., 2016).

According to Thar and Kuhl (2004), ultra-weak biophotons can be guided along a mitochondrial and microtubule network that can act as optical waveguides in neurons. So the protein-protein biophotonic interactions and mitochondrial interaction networks can constitute the neural biophotonic communication network. In a recent article (Burgio et al., 2020) we proposed that the interaction between the tubulins quantum computer ($QC_T$) in the cytoskeleton, where the subscript "T" stands for "tubulins", and the genome (DNA plus epigenome) is mediated by the biophotons, which are the quanta of the Genomic Quantum Electromagnetic Field ($QEF_G$).

The "orchestrating" (or coherent) action of the biophotons allows the orchestrated objective reduction (Orch-Or) of the tubulin microtubules in the cytoskeleton. This means that Orch-Or could be genome-induced. A beam of biophotons emitted at the B site of the epigenome could be transmitted to the brain even several centimeters away.

In the brain of a schizophrenic subject, the flow of biophotons reaching the microtubules should be so low that it cannot induce decoherence to all the overlapping tubulins, but only to a very few of them. Therefore, the tubulins persist mainly in a state of quantum superposition and the schizophrenic mind remains "trapped" in a quantum computational mode, corresponding to a permanent unconscious state.

In contrast, the biophoton flux in the autistic brain is so high that tubulin dimers persist mainly in a classical state, and the autistic mind is "trapped" in a classical computational mode, corresponding to a permanent conscious state. Perhaps the most striking examples are some genetic modifications common to autism and schizophrenia and particularly the Copy Number Variants (CNVs) that are formed in the same genomic sites, but in opposite forms, for example duplication in the autism (Clements et al., 2017) and deletion in schizophrenia (Van et al., 2017) in the same 22q11.2 region: therefore more probably effects than causes of the two disorders. Also in this case we are facing a reversal, or rather a straightening of the dominant model.

We believe that new methods can be developed to increase / decrease biophotonic activities in neurons in order to reverse the abnormal biophotonic fluxes in both autism and schizophrenia, with the hope that patients can improve their mental condition. High levels of reactive oxygen species (ROS), which cause oxidative stress, are present in autistic patients. Pangrazzi et al. (2020) described the major alterations in the expression of genes coding for enzymes involved in the ROS scavenging system, in autistic patients. Numerous drugs have been described capable of decreasing reactive oxygen species (ROS scavengers) and consequently a reduction of biophotons useful in autism could be obtained.

In a recent paper Wang et al. (2016) stated that biophotonic activities and transmission dominate the information neural processing and encoding mechanism in the brain, then biophoton spectral redshift could improve and strengthen cognitive abilities.

Sun et al. (2010) found that different stimulation of spectral light (infrared, red, yellow, blue, green and white) at one end of the spinal sensory or motor nerve roots resulted in a significant increase in biophotonic activity.

Since an increase in ROS species is not a viable strategy, we suggest the use of Brain-Computer Interfaces generating external light stimuli, for the possible treatment of schizophrenic patients. Biophotonic methods for brain-computer interfaces have already been described (Soraghan et al., 2007).

The review by Martins et al. (2020) describes the state of the art of human brain / cloud interfaces by introducing the "cyborgization of Homo Sapiens" in which future cyborgs are wirelessly interconnected and individualism is suppressed for the benefit of the "collective" (see Borg from Star Trek).

They conclude that it is conceivable that within the next 20–30 years, neuronanorobotics could be developed to enable a safe, secure, instant, real-time interface between the human brain and biological and non-biological computer systems, by enhancing brain interfaces (BTBI), brain-computer interfaces (BCI) and, in particular, sophisticated brain/cloud (B/CI) interfaces. Such human B/CI systems can dramatically alter human/machine communications, promising significant human cognitive enhancement (Kurzweil, 2014; Swan, 2016).

## 9 Eliano Pessa, The Man, The Scientist, The Friend

I Massimo Pregnolato (MP) met Paola Zizzi and Eliano Pessa in the first months of 2007, shortly after having founded the QuantumBioNet.org network. My acquaintance with physicists was just started and over the years to come it will give me a lot, from many points of view. However, the acquaintance of Paola and Eliano soon turned into friendship. Eliano taught General Psychology in my same University, but I wasn't so lucky to met him before. Time later did I understand why a physicist of such great calibre ended up teaching psychology and not physics.

Our relationships began on a basis of scientific collaboration but we soon realized how many interests and points of view on life we had in common.

Those were the basis for a strong and lasting friendship that lasted over time until the moment of his untimely death. In the years to come we have collaborated both in research and teaching, but also in the planning and implementation of several Quantumbionet Workshops and in the growth of the network itself.

We often met even just for a light snack at lunch and chatted about everything. He had a deep understanding not only of physics but also of the affiliations and genealogies of various physicists. Eliano was an expert on complex systems, emergence and neural networks and together with Paola, we shared our interest in quantum consciousness, both human and animal. One of his weird interest concerned not only the possibility that the mind could act on matter, but the opposite, that the matter (therefore electromagnetic waves) could affect the human mind. This would have terrible implications if the waves were to be used for mind control, but at the same time, useful in helping the communication of people with sensory deficits.

Eliano was a passionate climber, for many years he was able to take a long work break to tackle expeditions and climbs the Andes and the Himalayas.

On his return his stories were always passionate and compelling. Punctually, a few months late, I received postcards from him from the places explored. His travels have also been a source of cultural enrichment for him.

One day he told me about Milarepa and his legends. Biology and culture, consciousness and world, subject and object, interior and exterior have continuity and find, in the "creative transcendence" of consciousness and its experiences, a privileged degree of understanding. Together we have formulated a plausible hypothesis about the existence of different levels of consciousness in humans and animals.

He suggested that consciousness persists even in the face of minimal conditions, perhaps even in traumatic brain injuries. I agree that such a suggestion was justified at the biomolecular level through introduction of the hypothesis that Schrödinger proteins (i.e. tubulins) are the biological interface from quantum to classical computation, underlying quantum/classical consciousness processes and at the crossroad of memory and learning capacities. Eliano participated to the first Quantumbionet Workshop (Pavia, May 25th 2007) with the lecture: Problems in Theory of Phase Transitions in Biological Systems. To the second QBN Workshop organized by Paola (Padova, October 10th, 2008) with: Lorentzian vs Einstinien

Quantum Mechanics and The Role of Environment. To the third QBN Workshop (Pavia, September 24th 2010) with the lecture entitled: Quantum Networks.

On the day of April 27, 2013 a core international group of investigators, offering expertise in the fields of psychiatry, biochemistry, physics, computational neuroscience, mathematics, philosophy and theology, gathered in Palermo, Sicily under the auspices of the global Quantum Paradigm of Psychopathology (QPP) initiative with the aim of assessing the potential relevance of quantum physics and quantum chemistry to the mapping of mind-brain relations in normal and abnormal states of consciousness applicable to humans and non-human animals.

The QPP conference in Palermo has marked a definite turning point in the foundational perspective of many of the group's participants regarding the study of psychopathology, particularly mood disorders. One reason for this turning point stems from a realization that two of the most common forms of psychopathology, major depression and bipolar disorder, may be recognizable by means of biomolecular markers. Long years of theoretical study by independent investigators have finally culminated in a convergence of their insights via quantum paradigms that now promise to illuminate, through the empirically tangible route of such new biomolecular markers, pathological phenomena of the conscious brain, thus potentially both confirming in fact and further harmonizing the diverse prior contributions of these conceptually innovative psychiatrists, biochemists, molecular biologists, philosophers and theologians. Massimo Cocchi, Lucio Tonello, Fabio Gabrielli, Massimo Pregnolato, Paola Zizzi, Eliano Pessa, and their collaborators have forged links between serotonin and quantum phenomena via membrane biophysics in depression and psychosis. Even the absence of highly complex synaptic connections among neurons does not preclude the presence of at least rudimentary phenomenal experience in organisms endowed with superposed microtubular dimers, ordered water, membrane ion channels, and/or crucial lipid raft assemblies connected to selected second messenger systems. In addition, quantum-biophysical aspects of these and/or other yet unmapped structures and related processes may prove to be potent factors in the deeper etiologies and improved treatments of psychiatric disorders. To these assumptions Eliano contributed with his seminal lecture: "*Towards an integrated model of cytoskeletal quantum dynamics*". It seems to be consistent the hypothesis that Schrödinger proteins interactoma and in particular the cytoskeleton nanowire network is the best biological interface for potential expression of consciousness, being typical and specific for each animal species and that consciousness is always a potential. It's very fascinating to think that every animal possess a primary Schrödinger proteins complex (cytoskeleton) and even in the absence of circulating serotonin there is a potential of consciousness that is essential to the behavior of some life forms, while other species such as invertebrates, procariotes and even archea possess expertise in their own domain probably mediated by their own Schrödinger proteins interactoma (Cocchi et al., 2011).

I Paola Zizzi (PZ) was introduced to Eliano Pessa by Massimo Pregnolato at the first QuantumBionet workshop (Pavia, May 25th 2007) organized by Massimo.

All three of us became friends and collaborators. When I met Eliano, I had recently started writing my PhD thesis in logic on quantum metalanguage. Eliano was very interested in this topic, also because he saw analogies with non-unitary operators in quantum field theory. Giuseppe Vitiello (Peppino) with whom Eliano had collaborated had the same impression. I think that Eliano's closest and most fruitful collaboration in QFT was that with Peppino, with whom he shared his philosophical approach to theoretical physics.

Many years have passed since that beginning and yet everything still seems so alive to me ... many memories of conferences, congresses, workshops and meetings where Eliano, Massimo and I went together.

I particularly remember the Third QPP Meeting held in Bologna on 19–20 June 2014 (just 1 year after the "Palermo Declaration") where I, Eliano, Max and Peppino continued, even after the conference, with many discussions, intense correspondence, endless phone calls and e-mails.

Anyone who has had a close collaborator and friend knows well how beautiful, exciting and vital for our intellect all this, and how much loneliness, how empty it feels when this friend is no longer there. It is the feeling of the orphan, as a friend of mine said when his closest friend and collaborator of him died.

Eliano behaved towards his friends, colleagues and students with kindness and generosity. He liked to take care of others, he helped everyone, with advice, mountains of bibliographies (as many as he gave me), with lent books, articles, endless discussions. A lovable and cultured person, easy-going, who has left a terrible void in my life and in that of all his friends and colleagues.

During the 11 years of our collaboration, Eliano and I worked together on some different topics such as QFT, quantum computing, psychopathology (particularly mood disorders), and artificial intelligence (AI).

I remember many of our discussions about QFT, sometimes arguing, because Eliano didn't "believe" in elementary particles while I did.

But yet, as a young man Eliano met Bruno Touschek, a Austrian physicist famous for research on particle accelerators, of which he was one of the pioneers particularly during his Italian period in Frascati.

Bruno, being a Jew on his mother's side, had been persecuted. Eliano considered Bruno to be a Master, and told me that he suffered the pains of hell when Bruno died relatively young in 1978 (the very year in which I graduated).

Eliano said that Bruno had also taught him something much more important than any formula: to always be himself and dignified in all circumstances. And in fact Eliano certainly did not lack dignity.

One of the papers that Eliano and I wrote together in QFT was "From SU (2) Gauge Theory to Qubits on the Fuzzy Sphere" in in 2014. I remember that while we were writing the paper, Eliano often repeated that 1 day we should write a book entitled "The mysteries of SU (2)". That book was never written, and now Eliano is gone.

In addition to quantum field theory, Eliano had a great interest in complex systems and artificial intelligence (AI), and was a great expert in neural networks.

He and I wrote a paper together in 2009 on AI at the quantum level. The paper was entitled “Brain-Computer Interfaces and Quantum Robots”.

## 10 Conclusions

In this paper in memory of Eliano Pessa we wanted to recall the highlights of our cooperation and friendship, pointing out the human qualities of the person even before those of the scientist. We wanted to remember him through the themes most dear to him that involved us in research collaborations. We talked about the definition of “mind” in terms of logical/metalogical modalities, ie classical/quantum logic for ordinary thinking and classical/quantum metalanguage for meta-thought. We have introduced the concepts of metalanguage and object language, their relationships and differences and we have shown that the axiom of identity belongs to the metalanguage, unlike the law of identity, which belongs to the language of the object.

With Eliano we came to the conclusion that a robot can never acquire self-awareness, demonstrating that while in a classical metalanguage the axiom of identity is absolute, in a quantum metalanguage it is probabilistic and above all it is in the quantum case, where the “ non-algorithmic side “of the human mind seems to be the insurmountable boundary for a robot.

What we call the “pillars” of the human mind, which distinguish it from a robot, are: Tarski’s truth predicate, the axiom of identity and the rule of cut, all three belonging to the metalanguage. We have suggested that, in the quantum case, during the programming process, the programmer’s mind and the quantum robot might get entangled and we have compiled some recent discoveries in quantum epigenetics and related them to a new non-invasive approach to the Brain-Computer-Interface based on quantum metalanguage and on a theoretical architecture of quantum cyborgs.

These pioneering concepts will take a few decades of work to implement, but like all frontier technologies they could lend themselves to doing good for humanity while at the same time being misused for the purposes of domination and power. The hope is that good will prevail, and in our opinion this is possible only if humanity manages to make the best use of the gift of metalanguage, where its greatness lies. And by greatness we mean understanding, empathy and sharing, all qualities that machines do not have. People who use metalanguage little, let themselves be dominated by reason and logic, On the other hand, people who don’t know they have a metalanguage or don’t know how to use it are confused and at the mercy of events. Finally, people who use metalanguage for evil have realized, even without knowing where it comes from, that they have a power, which can be used above all on the second category, that of confused humanity. Therefore, metalanguage, while being the highest way of thinking, does not deprive us of free will, rather it is what gives it to us. It is not up to us to choose the destiny of man but we cannot in any case stop a scientific progress that proceeds unstoppable also through the intuition of many researchers and pioneers such as Eliano Pessa was.

## References

Benioff, P. (1998). Quantum robots and environments. *Physical Review A, 58*, 893–904.

Burgio E. Lucangeli D., Zizzi P. (2020) *A model of quantum epigenetics in neuropsychiatry*. https://www.researchgate.net/publication/339841651. To appear in quantum Psyche-2, Federico Carminati and Giuliana Galli Carminati Eds.

Clements, C. C., Wenger, T. L., Zoltowski, A. R., Bertollo, J. R., Miller, J. S., de Marchena, A. B., Mitteer, L. M., Carey, J. C., Yerys, B. E., Zackai, E. H., Emanuel, B. S., McDonald-McGinn, D. M., & Schultz, R. T. (2017). Critical region within 22q11.2 linked to higher rate of autism spectrum disorder. *Molecular Autism, 8*, 58.

Cocchi, M., Tonello, L., Gabrielli, F., Pregnolato, M., & Pessa, E. (2011). Quantum human & animal consciousness: A concept embracing philosophy, quantitative molecular biology & mathematics. *Journal of Consciousness Exploration & Research., 2*(4), 547–574.

Gentzen, G. (1969). In M. E. Szabo (Ed.), *The collected papers of Gerhrard Gentzen* (pp. 68–131). North-Holland.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik, 38*, 173–198.

Hameroff, S., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness? In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness—the first Tucson discussions and debates* (pp. 507–540). MIT.

Kumar, S., Boone, K., Tuszynski, J., Barclay, P., & Simon, C. (2016). Possible existence of optical communication channels in the brain. *Sci Rep 6*, 36508. https://doi.org/10.1038/srep36508

Kurzweil, R. (2014). *Get ready for hybrid thinking*. Ted Talk. https://www.ted.com/talks/ray_kurzweil_get_ready_for_hybrid_thinking/transcript

Martins, N. R., Angelica, A., Chakravarthy, K., Svidinenko, Y., Boehm, F. J., Opris, I., Lebedev, M. A., Swan, M., Garan, S. A., Rosenfeld, J. V., & Hogg, T. (2020). Human brain/cloud interface. *Frontiers in Neuroscience, 13*, 1–24.

Pangrazzi, L., Balasco, L., & Bozzi, Y. (2020). Oxidative stress and immune system dysfunction in autism Spectrum disorders. *International Journal of Molecular Sciences, 21*, 3288.

Pati, A. K., & Braunstein, S. L. (2008). Can arbitrary quantum systems undergo self-replication? In D. Abbott, P. C. V. Davies, & A. K. Pati (Eds.), *Quantum aspects of life* (pp. 223–231). Imperial College Press.

Penrose, R. (1989). Shadows of the mind: A search for the missing science of consciousness. In *The Emperor's New Mind: Concerning computers, minds and the laws of physics*. Oxford University Press.

Pessa E., Zizzi P. (2009) Brain-computer interfaces and quantum robots. In *Contributed paper to the conference "mERGERs". Physical and cognitive mutations in humans and machines, Laval (France)*, 24–25 April 2009. https://arxiv.org/abs/0909.1508

Sambin, G. (2007). *Per istruire un robot*. Libreria Progetto.

Sambin, G., Battilotti, G., & Faggian, C. (2000). Basic logic: Reflection; symmetry, visibility. *The Journal of Symbolic Logic, 65*, 979–1013.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–458.

Soraghan, C., Matthews, F., Markham, C., Barak, P. A., & Tomas, E. W. (2007). Biophotonic methods for brain-computer interfaces. In *Photonics Ireland 2007, September 24th–26th, 2007*. (Unpublished).

Sun, Y., Wang, C., & Dai, J. (2010). Biophotons as neural communication signals demonstrated by in situ biophoton autography. *Photochemical & Photobiological Sciences, 9*(3), 315–322.

Swan, M. (2016). The future of brain-computer interfaces: Blockchaining your way into a cloudmind. *Journal of Evolution and Technology, 26*, 60–81.

Tarski, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research, 4*, 13–47.

Thar, T. R., & Kuhl, M. (2004). Propagation of electromagnetic radiation in mitochondria? *Journal of Theoretical Biology, 2004*(230), 261–270.

Turing A. (1948) *Intelligent machinery.*https://weightagnostic.github.io/papers/turing1948.pdf

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Umezawa, H. (1993). *Advanced field theory: Micro macro and thermal concepts*. American Institute of Physics.

Umezawa, H., & Vitello, G. (1985). *Quantum mechanics*. Bibliopolis.

Van, L., Boot, E., & Bassett, A. S. (2017). *Update on the 22q11.2 deletion syndrome and its relevance to schizophrenia*. *Current Opinion in Psychiatry, 30*(3), 191–196.

Vitiello, G. (1974). *Dynamical rearrangement of symmetry*. Ph.D. Thesis. The University of Wisconsin. Dissertation Abstracts International 36–02, B: 0769. http://inspirehep.net/record/101641/file.

Vitiello, G. (1995). Dissipation and memory capacity in the quantum brain model. *International Journal of Modern Physics B, 9*, 973–989.

Wang, Z., Wanga, N., Lia, Z., Xiaoa, F., & Daia, J. (2016). Human high intelligence is involved in spectral redshift of biophotonic activities in the brain. *PNAS, 113*(31), 8753–8758.

Wheeler, A. (1962). *Geometrodynamics*. AC Press.

Zadeh, L. A. (1996). *Fuzzy sets, fuzzy logic, fuzzy systems*. World Scientific Press.

Zizzi, P. (2001). Quantum computation toward quantum gravity. *General Relativity and Gravitation, 33*(8), 1305–1318.

Zizzi, P. (2005). Qubits and quantum spaces. *International Journal of Quantum Information, 3*(1), 287.

Zizzi, P. (2008). Quantum robot: Quantum mind control on a quantum computer. *Journal of Physics: Conference Series, 67*, 012045.

Zizzi, P. (2010). *From quantum metalanguage to the logic of qubits.* PhD Thesis, arXiv:1003.5976.

Zizzi, P. (2011a). Quantum mind from a classical field theory of the brain. *Journal of Cosmology, 14*, 4472–4483.

Zizzi, P. (2011b). The non-algorithmic side of the mind. In *Workbook of the 4th Quantumbionet Workshop, the dawn of quantum biology, Crema, 30th September 2011* (p. 9) https://arxiv.org/abs/1205.1820

Zizzi, P. (2018). *The disintegrated self. Lectures given at Science of Consciousness Research Group*. Department of General Psychology Padua University.

Zizzi, P. (2020a). *Quantum field theory as quantum metalanguage*. Work in progress.

Zizzi, P. (2020b). *Quantum information hidden in quantum fields*. *Quantum Reports, 2*(3), 459–488. https://doi.org/10.3390/quantum2030033

Zizzi, P. (2020c). Quantum minds and quantum computers: A subtle border. *In Invited talk at the conference: "Convegno Nazionale di biologia Quantistica", Urbino (Italy) 6–7 June 2020*. https://www.researchgate.net/publication/3420046

Zizzi, P. (2020d). *Children, robots and metalanguage*. https://www.researchgate.net/publication/344227933

Zizzi, P., & Pregnolato, M. (2012a). Looking for the physical, logical, and computational roots of the mind. *Journal of Consciousness Exploration & Research, 3*(4), 425–431.

Zizzi, P., & Pregnolato, M. (2012b). Quantum logic of the unconscious and schizophrenia. *Neuroquantology, 10*(3), 566–579.

Zizzi, P., & Pregnolato, M. (2020). Quantum computing and the quantum mind: A new approach to quantum gravity. In L. M. Caligiuri (Ed.), *Frontiers in quantum computing* (pp. 153–175). Nova Science.

# Index

G. Minati (ed.), *Multiplicity and Interdisciplinarity*, Contemporary Systems Thinking, https://doi.org/10.1007/978-3-030-71877-0