Integrated Science 5

Nima Rezaei *Editor*

Transdisciplinarity

Springer

Integrated Science

Volume 5

The **Integrated Science** Series aims to publish the most relevant and novel research in all areas of Formal Sciences, Physical and Chemical Sciences, Biological Sciences, Medical Sciences, and Social Sciences. We are especially focused on the research involving the integration of two of more academic fields offering an innovative view, which is one of the main focuses of Universal Scientific Education and Research Network (USERN), science without borders.

Integrated Science is committed to upholding the integrity of the scientific record and will follow the Committee on Publication Ethics (COPE) guidelines on how to deal with potential acts of misconduct and correcting the literature.

Nima Rezaei
Editor

Transdisciplinarity

Springer

Editor
Nima Rezaei
Universal Scientific Education
and Research Network (USERN)
Stockholm, Sweden

ISSN 2662-9461 ISSN 2662-947X (electronic)
Integrated Science
ISBN 978-3-030-94650-0 ISBN 978-3-030-94651-7 (eBook)
https://doi.org/10.1007/978-3-030-94651-7

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

This book series would not have been possible without the continuous encouragement of my family. I dedicate this book series to my daughters, Ariana and Arnika, hoping that integrated science could solve complex problems and make a brighter future for the next generation.

Preface

During all these years, starting as a junior medical student and researcher, and then as a faculty member, I often wondered, what makes a perfect researcher, a perfect teacher, a mentor, or a leader? Realizing that none of these would make me a perfect human was one of the most important discoveries of my life.

As one dives deep into his own scientific field, interacting with his colleagues, and joining educational and research groups, we realize how high we've built the walls around ourselves, and those who share the same interests as us. We learn not to resist invaders into our territories and to think and behave as affiliates of a certain virtue.

Medicine, biology, mathematics, chemistry, physics, astronomy, and art are all parts of the heritage of ancient, true pioneers of knowledge. The enormity of this prodigious legacy can only reach its true potential when these segments reunite as a whole and into knowledge without borders. It is undeniable that science today is unintentionally mistaken for a line to draw boundaries with, a weapon to display power, or a rule to rank orders. We believe that the golden key to this reconciliation is by the hands of the scientists themselves, by the hands of artists, and by the hands of anyone who has an ability to share what they know, for the greater good.

The poem of "The Elephant in The Dark", which was written by Mevlana (Rumi), a Persian poet seven century ago, was originated in the ancient Indian subcontinent, when some Hindus bring an elephant to be exhibited in a dark room. This story is as an example of the limits of individual perception, while I believe that it could also emphasize the important role of transdisciplinary. Transdisciplinary is a problem-oriented strategy that systematically integrates knowledge and experience from various scientific and social partners to solve real-life situations. *Transdisciplinary* is an effective strategy that interconnects the scientific and abstract thinking theories with empirical, experimental findings to address prominent universal disagreements, complex social, economic, public health, environmental, and humanity issues such as poverty, sustainability, public health, equality, justice, and education. This is what the transdisciplinary approach provides us, by annealing and remodeling fragmented theories, methods, experiences, and information from disciplines and reconstructing a united comprehensive knowledge that is able to explain the wicked problem as well as creating novel, brilliant, and innovative approaches and solutions on the planetary scale. Such an approach should be willing to accept the contribution of various disciplines, cultures, and society.

This is what we are looking for it by establishment of the Universal Scientific Education and Research Network (USERN) with the main purpose of peaceful and humanitarian promotion of education and research, universally. It comprises more than 17,000 members from more than 120 countries in all scientific fields, where more than 600 top 1% scientists from different disciplines and 19 Nobel/Abel Laureates as the advisory board members who would manage and supervise the educational and research programs in different disciplines. It is to be hoped that scientific discussion of scientists out of this dark room provides a better concept of the whole elephant, which consequently might help us to make the world as a better place for living for next generation.

Stockholm, Sweden Nima Rezaei MD, Ph.D.

Acknowledgment

I would like to express my gratitude to the Editorial Assistant of this volume of Integrated Science book series, Dr. Simin Seyedpour. The book would not have been completed without her contribution.

Contents

1 Introduction to Integrated Science: Transdisciplinarity

Nima Rezaei and Simin Seyedpour

> *"Leave that nothing that looks like it's something;*
> *Seek that something that looks like it's nothing".*
>
> Molana Jalaleddin Rumi

Summary

As the human being started his journey to discover the world, he tried to comprehend the universe by categorizing knowledge into various disciplines, which leads to ending up in the constant expansion of specialization. Although scientific research enormously revolutionized human insight toward his surrounding environment, it remained insufficient to address complex problems. *Transdisciplinary* is a problem-oriented strategy that systematically integrates knowledge and experience from various scientific and social partners to solve real-life situations.

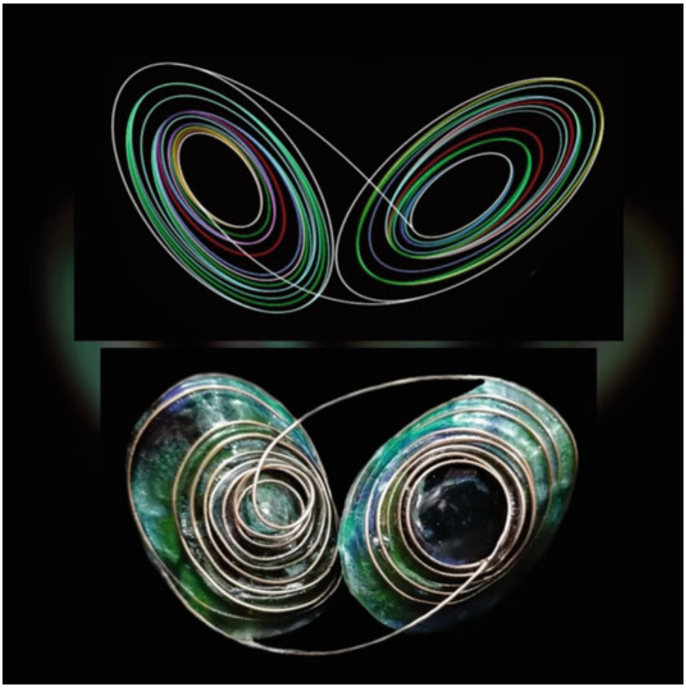
N. Rezaei (✉) · S. Seyedpour
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Tehran, Iran
e-mail: rezaei_nima@tums.ac.ir

N. Rezaei · S. Seyedpour
Research Center for Immunodeficiencies, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

N. Rezaei
Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

S. Seyedpour
Nanomedicine Research Association (NRA), Universal Scientific Education and Research Network (USERN), Tehran, Iran

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_1



The Butterfly effect Physics science and handicrafts [Adapted with permission from the Association of Science and Art (ASA), USERN; Artist: Sepideh Sargoli, Method: resin art and brass].

Butterfly effect is the name of a phenomenon that occurs due to the sensitivity of turbulent systems to the initial conditions. The propeller effect means that a slight change in initial conditions can lead to far-reaching and unpredictable results in system outputs, and this is the cornerstone of chaos theory. In the theory of chaos or disorder, it is believed that in all phenomena, there are points at which a small change will cause great changes.

The code of this chapter is 01001001 01101110 01110100 01100101 01100111 01110010 01100001 01110100 0110100101101111 01101110.

1 Introduction

As the human being started his journey throughout the universe, He tried to understand the world by categorizing his perceptions into various scientific fields named as disciplines. Science divided nature and society into separate divisions such as formal sciences, natural sciences, social sciences, etc. Although academic disciplines enormously revolutionized human civilization, survival, and insight toward his surrounding environment, they caused constant expansion of specialization [1].

University faculties, discipline-specific congress, subspecialized seminars, peer-review journals with specific scientific scopes are the by-products of knowledge classification that limited individuals' exposer to state-of-the-art knowledge produced within the boundaries of a single discipline [1, 2].

However, despite the tremendous advancement of disciplinary knowledge, the world faces deep, complex, and fundamental challenges that remained to be solved by traditional research and disciplinary expertise. In Volume 1 of the integrated science book series, authors discussed complex problems and various complex problem-solving strategies, including interdisciplinary, multidisciplinary, and transdisciplinary approaches. Each method had its unique characteristics and applications.

The multidisciplinary approach is efficiently applicable in settings that require the parallel or complementary collaboration of several specialties. For example, various medical specialists should collaborate and consult with one another to fulfill a common goal, patient's treatment. Each specialist will participate in part of the problem-solving process using their knowledge in the boundaries of their disciplinarily. It is an additive strategy that adds methods, information, and solution from various disciplines for solving a common problem [3].

The interdisciplinary method is designed for situations where there is a need to integrate two or more disciplines. Unlike multidisciplinary, the researchers need to interact with other partners to accomplish and enrich their knowledge with ideas and methodologies from different areas, so the boundaries among academic disciplines will become transparent. This integration can be among similar disciplines such as biomedical sciences, resulting from biology and medicine integration, or very distinct disciplines such as medical ethics, resulting from social science and medicine integration [4].

The various knowledge integrating methods had no propriety over each other. They should be considered complementary to each other and are reflecting the limited access of human beings to what we can name as the "reality."

2 Transdisciplinary, a Strong Knowledge Integration Method

Global disasters such as the COVID-19 pandemic induces multidimensional challenges on universal scale. A single or even multiple disciplines remained insufficient to respond to such real-life problems and require a higher level of knowledge integration provided by the third knowledge production strategy, the transdisciplinary method [5].

Transdisciplinary is a problem-oriented method that combines, integrates, and synthesize knowledge across disciplinary boundaries to generate a comprehensive solution. Transdisciplinary research focuses on a complex problem that cannot be solved by methods, concepts, theories, and knowledge produced by a single or even multiple disciplines. It requires another fundamental component, society [6].

The transdisciplinary researchers are doing science with society rather than implementing science on or for the community. It considers and unifies diverse points of view from experts, researchers, scientists, stakeholders, citizens, people, and public and private sectors. These partners participate in equal weight throughout the whole process [7].

In the context of social issues, the knowledge is not only limited to academic professionals. The governmental and non-governmental organizations, stockholders, social actors, and any other participants are needed to overcome the complexity of issues regarding the common good of society [8–10].

Transdisciplinary is an effective strategy that interconnects the scientific and abstract thinking theories with empirical, experimental findings to address prominent universal disagreements, complex social, economic, public health, environmental, and humanity issues such as poverty, sustainability, public health, equality, justice, and education. Some of the most well-known examples of transdisciplinary knowledge production results are bioinformatics and nanomedicine, where informatics technologies or nanotechnology are implemented in the context of biomedical problems. Climate change and sustainability are other examples of transdisciplinary research applications which has been broadly described.

3 Transdisciplinary Versus Multi- and Interdisciplinary Method

Transdisciplinary is discriminated from other knowledge-producing models such as interdisciplinary and multidisciplinary in the following characteristics:

- **First**: In this approach, methodologies, theories, experiences, and intellectual concepts are transferred among various disciplines to get novel, innovative, and beneficial applications. One should consider the possible modifications which might happen or be requires while moving cross-disciplines.
- **Second**: The knowledge integration and unity are not only happening at the level of disciplinaries but also proceeding beyond scientific and academic disciplinaries and involves experts, public, government, stakeholders, and social actors.
- **Third**: The social actors, including people, stockholders, etc., are not the consumers of newly produced or technology. They are not observers, and their ideas, comments, and suggestions are not considered complementary to the result of integration emerge from an interdisciplinary method. They are actively involved and participate in equal weight to other scientific partners from the first step. They will define and formulate problems, determine challenges, limitations,

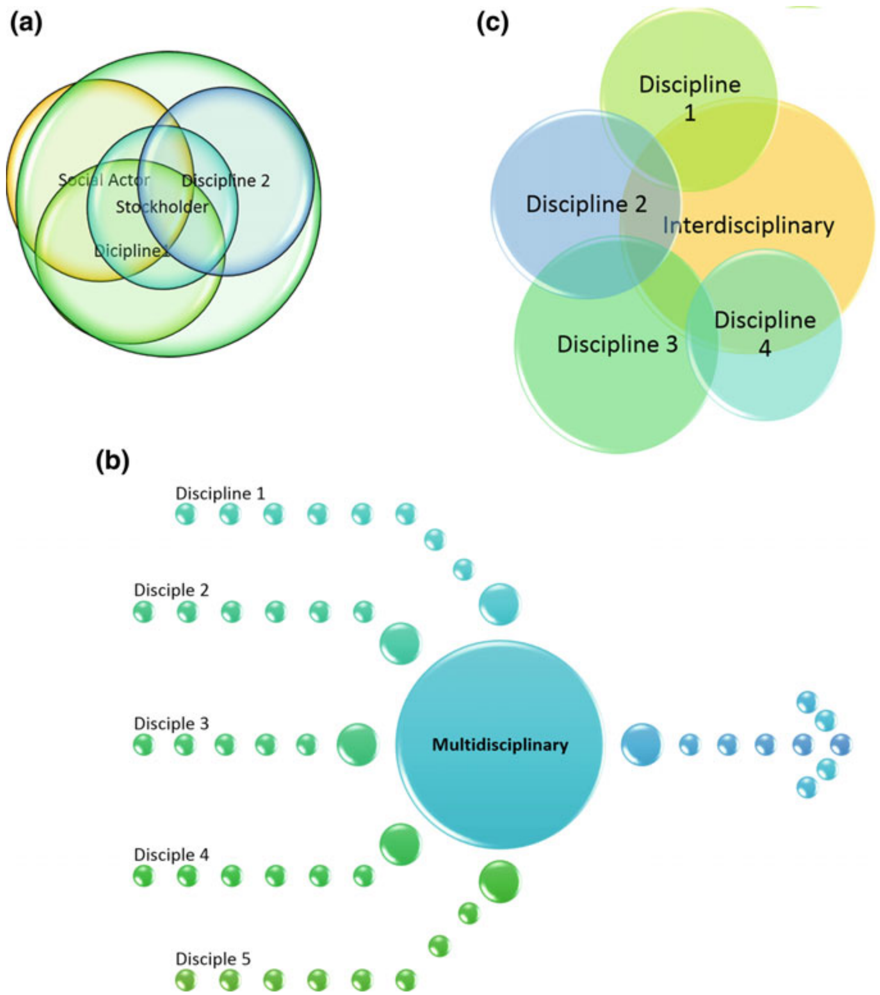


Fig. 1 Schematic manifestation of knowledge integration in transdisciplinary (Fig. 1a) versus multidisciplinary (Fig. 1b) versus interdisciplinary (Fig. 1c)

and obstacles they faced, and propose practical, innovative methods to overcome the complex problem. The scientist and researchers will explain theoretical models, rules, intellectual methodologies, while society focuses on their relevance, implications, and practical effectiveness in solving a real-life problem (Fig. 1) [7, 11].

4 Transdisciplinary Origin

Transdisciplinary is not a newly emerged concept. It seems the concept is released in 1970 at a seminar held at the University of Nice. The idea of transdisciplinary was emerged from a discussion on interdisciplinarity seminar by a Swiss psychologist named Jean Piaget. The term was used to suggest a more systematic and organized strategy for integrating methods, concepts, and ideas; as Piaget stated, transdisciplinary was defined as "*a system of knowing beyond disciplinary boundaries*". In line with Piaget, Jack Lee Mahan in the USA proposed a comparable idea of systematic integration of knowledge. He explained the transdisciplinary concept in his thesis entitled "*Toward Transdisciplinary Inquiry in the Humane Sciences*". He emphasized the importance of considering ethical issues while addressing a complex social problem. He defined transdisciplinary as an integrative strategy that breaks, exposes, and transcends disciplinary borders. Such integration leads to the creation of novel applicable intellectual concepts and theories for solving real-life problems [12].

5 Transdisciplinary Research Process

The question is how participants should interact to meet transdisciplinary research goals? What methods, steps, and strategies should be implemented? What skills are required to facilitate knowledge and know-how exchange among disciplines and involved individuals? Although transdisciplinary is a systemic, rational, and organized form of knowledge integration method, there is no unique methodology for conducting transdisciplinary research in the literature. To understand this, we will review some of the critical concepts in conducting the transdisciplinary research process.

The transdisciplinary research process has three main steps:

Problem formulation, data integration, and providing practical solutions.

5.1 Problem Formulation and Interpretation

Transdisciplinary research starts with a socially related problem and focuses on providing sources, knowledge, experiences to solve that problem. In other words, it deals with those complex problems which are affecting society and people. It will be the best method to address the real-life challenges that existing disciplines and approaches are failed to solve.

Problem framing is a critical step because it will determine what kind of knowledge or experiences are expected to be involved in the project?

The transdisciplinary method addresses questions that have a complex essence. It is not apparent what academic discipline, organization, or experiences are responsible for solving them. The existing strategies are insufficient to provide a comprehensive strategy that overcomes the conflicts among its participants, stakeholders, and involved individuals. In the beginning, the team may have no consensus on the problem. They may have no idea about its origin and what might cause it, how to approach it using the expertise and knowledge they gained in the borders of their academic disciplines, how to solve the problem using the strategies and methodologies in their field of expertise.

The transdisciplinary method allows all involved participants to comprehend various dimensions of the issue when addressing it and to explore the problem's complexity. To accomplish this aim, they should implement and interrelate scientific, theoretical points of view with real experiences and practical knowledge. Furthermore, they should accept the common good as the essential principle in their suggestions while reflecting their benefits as a participant [8, 13, 14].

The problem (e.g., sustainability, public health, poverty) will be the core component of a huge system in which various participants are going to be involved and contribute, governmental and non-governmental organizations, social and voluntary communities, companies, agencies, individuals, social actors, and stakeholders will collaborate and will closely interact with scientific participants, academics, experts from related disciplines and researchers in a systematic manner, that is they will interact with one another throughout the research process to solve the common problem. Furthermore, they may express their opinion from their perspective regarding the issue, introduce methodologies, suggestions, tools, and technologies that can be implemented to improve the situation.

The involved sectors should actively involve and collaborate throughout all three stages rather than contributing to a part of the process.

Participants will bring together the key concepts from their expertise to identify the problem. They will purpose methods, strategies approached, tools, technologies, measures for investigating its causality and origin, and assess the effectiveness of suggested interventions to solve the problem. Although participants can involve in all stages of transdisciplinary research, and their suggestions and ideas should potentially be considered to have the same value, the amount of each discipline or participants' contribution may differ related to the type of problem.

The scientific methods, theories, and suggestions are not priorities to practical, social, and real-life experiences provided by social partners. Instead, these partners should interact with one another and focus on taking advantage of the potential in various fields to produce a problem-oriented approach and knowledge strategy that will significantly change the problem [8, 13].

5.2 Data Integration Methods

Regarding the definition of a complex problem in which a single discipline is not adequate to comprehend the dimensions of the issue, the transdisciplinary approach invites various disciplines as well as partners to introduce and related the multiple factors related to initiation and maintenance of the problem, such as scientific, social, political, historical, technological, biological, mathematical, linguistic, physical, practical, economical, practical, etc. factors. Systemic thinking is supposed as one of the models which could be implemented during the transdisciplinary research process [11].

Systemic thinking mainly focuses on how different factors can influence one another. Transdisciplinary differs from interdisciplinary techniques in their method of integration. Concepts, data, methods, and theories are combining, attaching, and joining together among two or more disciplines while transdisciplinary research action via understanding the relationship among its various elements and interrelating, transforming, remodeling their functionality in a systematic manner to comprehend, solve, or prevent a complex real-life problem [8].

The result is not simply the quantitative or qualitative summation of data they produce but how these interactions are changing the functionality of each section to fulfill the system goal. For example, if you consider a computer as a system, collecting various parts of a computer in a box will not result in its functionality. However, the specific interaction among various components such as motherboard, keyboard, CPU, and mouse is essential (Fig. 1) [8].

Various integration methods have been purposed for transdisciplinary research [8]. For example, the integration can happen during a shared group learning or when one of the individuals or subgroups is selected to conduct integration. In deliberation among expert model, the social or scientific participants (entitled as experts) are working on using their expertise. Then a few experts are exchanged among groups which facilitates the integration.

In the second level of integration, the key concepts can be exchanges or redefined among various disciplines. New concepts can also be developed in this stage to address the new concepts among participants. The preexisting models, whether complex, soft models, can also be considered as an integrative method [8] (Fig. 2).

6 Transdisciplinary Outcomes

Transdisciplinary researchers aim to develop or produce a product such as tools, techniques, materials, software, exhibitions, etc. which can take part in problem-solving can also be used as a joint goal for participants to integrate their knowledge and expertise.

Another characteristic of transdisciplinary research methodology is implementing the abstract, scientific theories into real-world problems, examining their effectiveness, and providing feedback for remodeling, editing, or complementing those theories in the context of problems. So, it is a bidirectional interaction that constantly tests the newly produced or existing theories and practically implements them in the setting of the problem of interest [11, 13].

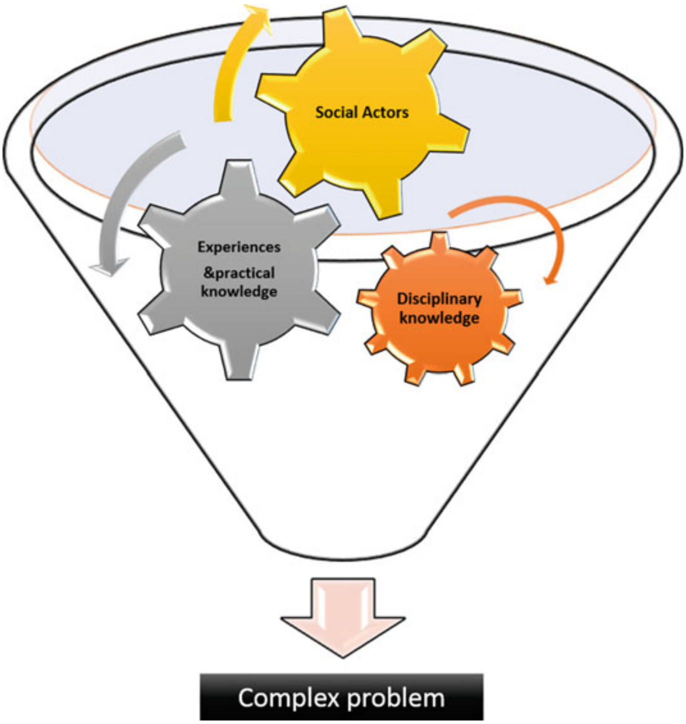


Fig. 2 Transdisciplinary research method implements different facilities provided by academic and non-academic actors and transform them in a systematic manner to reach the function of interest for solving a real-life problem

7 Transdisciplinary Pros and Cons

Transdisciplinary will facilitate new practical insights, theories, and ideas because it can bypass and solve the knowledge exchange barriers and conflicts among various participants. However, transdisciplinary cannot substitute the basic research methods. Instead, it is a form of applied research that requires the involvement of social and scientific organizations in determining challenges from the beginning. Similar to other polydisciplinary methods, transdisciplinary has its limitations and benefits. The proper knowledge transformation is a necessity of successful transdisciplinary research. Providing a common language for various partners' cooperation is a time-consuming process. It also requires communication and information synthesis skills for delivering a successful outcome. One of the challenges that participants face during transdisciplinary research is that they might not have the exact comprehension of information given by one another regarding the variety of perspectives and backgrounds about that problem. Therefore, we must consider this diversity and express and clarify concepts to avoid any misunderstanding [8, 15].

The key concepts used among a group of expertise may have various meanings from one discipline or community to another. Also, the participant may not be aware of the primary expressions and languages used in other disciplines or communities. Defining these key concepts at the beginning and using the no scientific expressions and daily language are supposed as solutions for this challenge.

Therefore, the participants should communicate obviously and understandably in a liberated atmosphere while adjusting their benefits in line with the common good [16].

In brief, a successful transdisciplinary research output has the following characteristics:

- It should become beneficial for both stockholders and researchers.
- It will facilitate the audience's access to the research results and outcomes.
- It provides a unique opportunity for participants to explore and learn together [8, 11].

8 Conclusion

It is time to let the human brain comprehend reality in the way it created to receive, by integrating thoughts, emotions, and external sensory output and release its unlimited creativity and imaginary potentials and cross the boundaries between disciplinarily to perceive the complex problems in its full dimensions.

This is what the transdisciplinary approach provides us, by annealing and remodeling fragmented theories, methods, experiences, and information from disciplines and reconstructing a united comprehensive knowledge that is able to explain the wicked problem as well as creating novel, brilliant, and innovative approaches and solutions on the planetary scale. Such an approach should be willing to accept the contribution of various disciplines, cultures, and society.

Core messages

- Transdisciplinary is a problem-oriented method that combines, integrates, and synthesize knowledge across disciplinary boundaries to solve a complex real-life problem.
- The knowledge is not only limited to academic professionals. Social actors are leading partners who will actively involve throughout all research processes.
- The participants should contribute actively and clearly in a liberated atmosphere while adjusting their benefits in line with the common good.

References

1. Casadevall A, Fang FC (2014) Specialized science. Infect Immun 82(4):1355–1360. https://doi.org/10.1128/IAI.01530-13
2. Fang FC, Casadevall A (2011) Retracted science and the retraction index. Infect Immun 79 (10):3855–3859. https://doi.org/10.1128/IAI.05661-11
3. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R (2003) Continuity of care: a multidisciplinary review. BMJ (Clin Res Ed) 327(7425):1219–1221. https://doi.org/10.1136/bmj.327.7425.1219
4. Jones C (2010) Interdisciplinary approach-advantages, disadvantages, and the future benefits of interdisciplinary studies. Essai 7(1):26
5. Westafer LM, Elia T, Medarametla V, Lagu T (2020) A transdisciplinary COVID-19 early respiratory intervention protocol: an implementation story. J Hosp Med 15(6):372
6. Arnold M (2013) Transdisciplinary research (transdisciplinarity). Encyclopedia of creativity, invention, innovation and entrepreneurship:1819–1828
7. Després C, Brais N, Avellan S (2004) Collaborative planning for retrofitting suburbs: transdisciplinarity and intersubjectivity in action. Futures 36(4):471–486
8. Pohl C, Hadorn GH (2008) Methodological challenges of transdisciplinary research. Nat Sci Soc 16(2):111–121
9. Hoffmann S (2016) Transdisciplinary knowledge integration within large research programs. GAIA-Ecol Perspect Sci Soc 25(3):201–203
10. Hoffmann S, Klein JT, Pohl C (2019) Linking transdisciplinary research projects with science and practice at large: introducing insights from knowledge utilization. Environ Sci Policy 102:36–42
11. Pohl C, Klein JT, Hoffmann S, Mitchell C, Fam D (2021) Conceptualising transdisciplinary integration as a multidimensional interactive process. Environ Sci Policy 118:18–26
12. Bernstein JH (2015) Transdisciplinarity: a review of its origins, development, and current issues
13. Pohl C, Hadorn GH (2007) Principles for designing transdisciplinary research. Oekom Munich
14. Pohl C, Perrig-Chiello P, Butz B, Hadorn GH, Joye D, Lawrence R, Nentwich M, Paulsen T, Rossini M, Truffer B (2011) Questions to evaluate inter-and transdisciplinary research proposals. TD-NET Network for Transdisciplinary Research
15. MacMynowski DP (2007) Pausing at the brink of interdisciplinarity: power and knowledge at the meeting of social and biophysical science. Ecol Soc 12(1)
16. Ezz I, FURLONG S, PAPAZAFEIROPOULOU A Large scale E_Government projects: the need for transdiciplinary collaborating teams. In: Articulo presentado en eGovernment Workshop, 2006. Citeseer

Nima Rezaei gained his medical degree (MD) from Tehran University of Medical Sciences (TUMS) in 2002 and subsequently obtained an MSc in Molecular and Genetic Medicine and a PhD in Clinical Immunology and Human Genetics from the University of Sheffield, UK. He also spent a short-term fellowship in Pediatric Clinical Immunology and Bone Marrow Transplantation in the Newcastle General Hospital. Since 2010, Dr. Rezaei has worked at the Department of Immunology and Biology, School of Medicine, TUMS; he is now the Full Professor and Vice Dean of International Affairs, School of Medicine, TUMS, and the Co-founder and Head of the Research Center for Immunodeficiencies. He is also the founding President of Universal Scientific Education and Research Network (USERN). He has edited more than 40 international books, has presented more than 600 lectures/posters in congresses/meetings, and has published more than 1000 articles in international scientific journals.

Simin Seyedpour gained her MD-MPH Diploma from the Tehran University of Medical Sciences in 2020. She does research on clinical immunology, cancer, and disease pathophysiology at the Research Center for Immunodeficiencies, Children's Medical Center, TUMS.

Chaos Theoretical Explanation to Each Development of Evolution Theory, Psychology, Physics, and Philosophy

2

Hideaki Yanagisawa

I hold time to be an order of successions.

Gottfried Wilhelm Leibniz

Summary

Three or more variables and continuous covariation are required to have a chaotic equation. Except for mathematical principles and historical facts, all-natural phenomena, including the existence and direction of time, obey chaos theory. Therefore, views on natural phenomena that do not consider the course of time must be corrected. Chaos theory comprises both a fixed state and a chaotic state and two different increasing and decreasing entropy directions. Entropy increases in all-natural phenomena except in evolution and a part of thinking. Total academic entropy increases if each academic field does not obey chaos theory. Each development of evolutionary theory, psychology, physics, and philosophy can be determined as myth, fixed (assertive) thinking, chaotic (non-assertive) thinking, and decreasing entropy thinking. Gene's learning evolution theory and Rogers' counseling are equivalent to reducing entropy thinking. Here, the Big Bang, dark matter, and dark energy theories do not obey chaos theory because the

The original version of this chapter was revised, which was inadvertently published prematurely before incorporation of the final corrections, which have now been updated.
The correction to this chapter can be found at https://doi.org/10.1007/978-3-030-94651-7_32

H. Yanagisawa (✉)
Marelli Corporation, 132 Shin-nakano, Ora-cho, Ora-gun, Gunma 370-0612, Japan

USERN, hideaki34, Ora-gun, Gunme, Japan

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_2

energy of the electromagnetic wave and the gravitational wave is considered no change during transmission. Therefore, a new gravitational equation considering the course of time is reported. A unifying style of thinking, considering time, may be equivalent to the new academic fields.

1 Introduction

Leibniz said, *"I hold time to be an order of successions."* However, Newton's absolute theory of time is more popular than Leibniz's idea with no equation. Here, all-natural phenomena obey chaos theory with continuous covariation. In chaos theory, Leibniz's idea is correct, and Newton's theory of time is wrong. In this book, a definition of time according to Leibniz's idea is reported with an equation. And all academic fields will be corrected with the new definition of time obeying chaos theory. Current academic disciplines and religion are divided into many fields.

Given that a subject's relation to another is pretermitted with analysis, abuse and war may arise. Research and academic disciplines without decreasing total entropy would harm living creatures. In contrast, a unifying style of thinking considering time is greatly beneficial to living creatures because of decreasing total entropy. It will become a standard in all academic fields, including each religion. By educating this mathematical standard, human thinking will be unified in the correct direction.

Therefore, abuse and war caused by separation will decrease in each field. Three variables are adequate in forming a chaos condition [1, 2]. The current cosmology accepts the Big Bang [3], dark matter [4], and dark energy theories [5] as correct. Such a condition does not allow us to draw a relation between the chaos condition and cosmology. However, chaos phenomenon can never exist without continuous covariation [6–13]. By this correction, theories on natural phenomena that do not consider the course of time must be corrected because, except for mathematical principles and historical facts, all-natural phenomena, including the existence and direction of time, obey chaos theory [12]. For example, the Big Bang, dark matter, and dark energy theories do not follow chaos theory because the electromagnetic wave (light) and the gravitational wave are considered no change during transmission. Therefore, a new gravitational equation considering the course of time was reported [12, 14, 15].

Theoretically, two modalities of thinking type can exist, a fixed type and a chaotic type. Thus, two are the directions with a change in thinking. One is direction increasing entropy, which destroys living creatures. The other is direction decreasing entropy, which is equivalent to evolution and a part of thinking [7–13, 16].

Entropy partially decreases for each academic field, while the entropy of academic fields is increasing as a whole because of specialization. Therefore, the common purpose of academia may be lost within each specialized field. In each academic field, the entropy change of the whole academia must be considered. Each development of evolutionary theory, psychology, physics, and philosophy can be determined as myth, fixed (assertive) thinking, chaotic (non-assertive) thinking, and decreasing entropy thinking. The theories of Copernicus, Darwin, Freud, Adler, Newton, Einstein, and Greek philosophers are equivalent to assertive thinking, while those of Imanishi, Jung, quantum mechanics, Russell, and Brouwer are equivalent to non-assertive thinking. Gene's learning evolution theory, Rogers' counseling, and the new gravitational equation are equal to decreasing entropy thinking.

Such a unifying style of thinking considering the course of time may be equivalent to the new academic fields.

2 Explanation of Chaos Theory

Here, we explain chaos theory, the relation between thinking and chaos theory, and some important preliminary results.

The contents of "Explanation of chaos theory" are similar to the author's articles [8, 11–13]. However, it is repeated in this report because of its importance.

2.1 Definition of Chaos Theory

The definition of chaos theory was reported as below [9]. Chaos theory can be defined as *"the qualitative study of unstable a periodic behavior in deterministic non-linear dynamical systems"* [17]. Chaos theory is a part of complexity theory that concerns itself with non-linear dynamic systems whose behavior does not follow clearly predictable and repeatable pathways. In linear systems, the relationship between an environmental factor and system behavior is predictable and easily modeled.

As the presence of an environmental factor increases, system behavior changes linearly in response to it. In contrast, behavior in chaotic systems might be perceived as unpredictable [18]. In this regard, such a chaotic state must not be confused with the term "random." In mathematical terms, "random" means the "statistics governed by or involving equal chances for each item" (New Oxford American Dictionary).

2.2 The Relation Between Continuous Covariation and Chaos Theory

Three or more variables and continuous covariation are required to have a chaotic equation [6]. In any chaos equation, fixed and chaotic solutions can be obtained that are continuous and have a bifurcation point between them, known as the Feigenbaum point [19].

For example, a chaos equation that is representative of chaos is expressed as follows:

$$Y(n+1) = p[1 - Y(n)]Y(n) \tag{1}$$

In Fig. 1, a schema near the Feigenbaum point is shown in parts *E*, *F*, and *G*, where the converging fixed (parts *C*, *D*, and *E*), localized (part *G*), and proliferating chaotic (part *H*) states are illustrated. The dotted line *F* is the Feigenbaum point. Except for mathematical principles and historical facts, all-natural phenomena obey chaos theory because of three or more variables and their continuous covariation between several phenomena, including matters and the mind.

3 The Relation Between Entropy Change and Chaos Theory

"Entropy" is a statistical word and was originally unrelated to any physical phenomena [20]. Entropy decreases when there is a change of direction from a chaotic state to a fixed state [6–13, 16], shown as the arrow *L* in Fig. 2. A schema of near

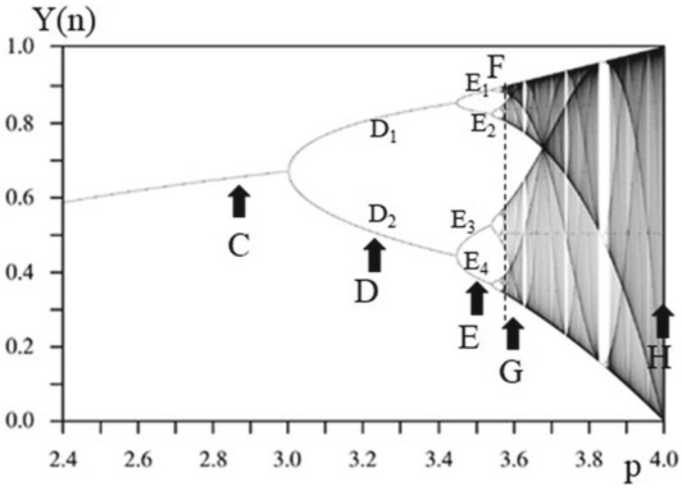


Fig. 1 Logistic map of Eq. 1. The converging fixed (parts *C*, *D*, and *E*), localized (part *G*), and proliferating chaotic (part *H*) states are illustrated

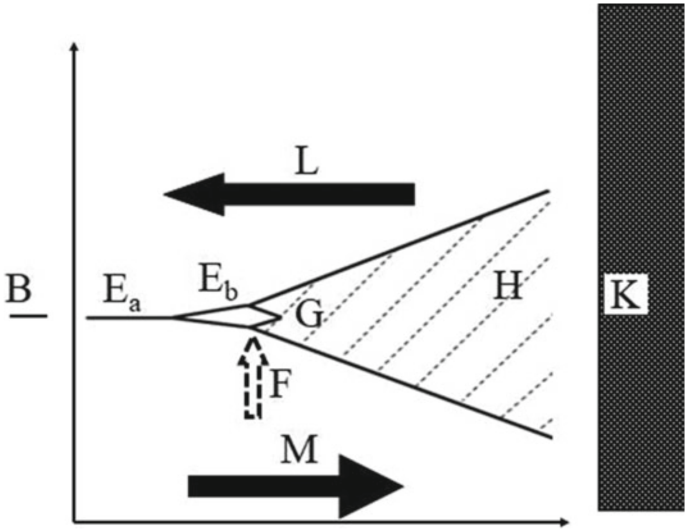


Fig. 2 Schema of complete fixed, incomplete fixed, chaotic, and random states. Each decreasing entropy and increasing entropy is shown as the arrows *L* and *W*

Feigenbaum point is shown as parts E_a, E_b, *G*, and *H* in Fig. 2. On the other hand, it increases whenever there is a change of direction from a fixed state to a chaotic state. This is shown as the arrow *M* in Fig. 2.

4 Mathematical Classification: Inside and Outside Chaos Theory

A chaos equation has either possible or impossible solutions. While impossible solutions are those with either no solution or with infinite solutions, possible solutions comprise complete fixed, incomplete fixed, chaotic, and random states [10–13, 16]. In complete fixed states, time is not required because no change occurs [21]. Examples are mathematical principles and historical facts, which do not change along with the environment. In chaos theory, a fixed state can become a chaotic state depending on the equation's variables, meaning that the state of a solution can also change as the environment changes. Therefore, in chaos theory, a fixed state is incomplete.

In Fig. 2, the extreme left side of parts E_a and E_b, (part *B*) is a complete fixed state and lies outside chaos theory. However, both the incomplete fixed (parts E_a and E_b) and the chaotic (part *G* and *H*) states are amenable to chaos theory. On the other hand, part *H* (part *K*) is a random state, not amenable to it. Since a chaos equation is based on mathematical principles, it is a complete fixed state, and it can be used to resolve incomplete fixed and chaotic states as well.

5 Relation Between Time and Chaos Theory

Isaac Newton [22] believed in the existence of absolute space and time despite all physical bodies disappearing in the universe. Conversely, Gottfried Wilhelm Leibniz [21] believed that space and time are logically and metaphysically related to physical bodies or events [23, 24]. Because Newton's time can never be scientifically proven, it is considered a myth.

Current science defines a second as the time it takes for an electromagnetic wave of a Krypton lamp to travel 299,792,458 m. Also, the length that the electromagnetic wave of cesium travels each second is set at 299,792,458 m. An absolute atomic clock presupposes the presence of absolute length, and an atomic absolute telemeter presupposes the presence of absolute time. Both definitions assume the constant speed of light; therefore, a "chicken and egg" contradiction exists in them. Since the relation between absolute distance and absolute time is contradictory in current science, a new definition of time is presented by this author [12, 25].

$$t = \frac{\log \frac{E(t)}{E(0)}}{k} \tag{2}$$

Here, t, $E(t)$, $E(0)$, and k are the time, the quantity of energy at time (t), the quantity of energy at time (0), and constant. Equation 2 was deduced from a stress equation.

$$\frac{\mathrm{d}E}{\mathrm{d}t} = kE \tag{3}$$

Here, E is the quantity of energy. This author considers that the definition of Leibniz's time is correct and that the existence and direction of time can be explained by chaos theory with continuous covariation [12].

6 The Relation Between Thinking and Chaos Theory

6.1 The Relation Between Entropy Decrease and Human Life

Through evolution and thinking, living creatures experience a decrease in entropy from a proliferating chaotic state (Part H) to a localized chaotic (Part G) or an incomplete fixed state (parts E_a and E_b) in Fig. 2.

Rearranging human thinking so that entropy decreases will lead to human satisfaction; moreover, humans may feel omnipotent when passing through the Feigenbaum point [19] (arrow F in Fig. 2). However, entropy does not decrease naturally in humans unless their thinking patterns are rearranged.

6.2 The Relation Between Counseling and Chaos Theory

It has been reported that the process of counseling equals that of solving scientific problems concerning chaos theory [9]. Here, the counselor, the client, and the information are equivalent to three or more variables and continuous covariation. During counseling, the counselor pays close attention to the client's thinking without referencing any objective standards. Because the counselor's focus is on the client's thinking pattern, their thinking must become the chaotic type. The counselor may not correct the client's ignorance regarding a particular objective standard; however, if the counselor merely repeats the client's expressions, then the counselor's thinking differs from the fixed type of thinking—merely parroting the client's expressions because of poor continuous covariation. Consequently, the counseling will fail because the client's thinking is unclear or because the client is confused.

During counseling, the counselor's primary skills involve listening closely, using reception, and synesthesia. The second skill allows counselors to confirm incomprehensible points by putting themselves in the clients' shoes. This confirmation process, which is equivalent to discovering a new theory or equation in science, is crucial to counseling and can be achieved professionally using a fixed type of thinking. Conversely, a chaotic type of thinking does not need to clarify incomprehensible points [9–13].

People who demonstrate a chaotic type of thinking cannot act autonomously due to dependence or lack of reference to objective standards. Therefore, merely listening closely, using reception and synesthesia, is (in and of themselves) insufficient to conduct counseling because, under such situations, there would be no decrease in entropy. This is further discussed in Sect. 7.4. Moreover, if entropy does not decrease, both the client and the counselor would never manage to achieve any form of lasting mental stability and, therefore, never be joyous.

6.3 The Relation Between Chaos Theory and computer's Human Face Recognition

Human face recognition by computers was developed by exchanging information between two computers [26, 27]. Here, the two computers and the information exchanged are equivalent to three or more variables with continuous covariation. Therefore, the relation obeys chaos theory, and entropy decreases because the two computers have a common purpose. It is equal to the counseling process and is similar to the dialectic process illustrated in Sect. 9.2 [12, 28].

7 The Relation Between Chaos Theory and Development of Evolution Theory

7.1 The Relation Between Evolution Theory and Myth

Before Darwin's theory of evolution was published and widely disseminated [29], it was believed that God had made all species. A new species is shown with arrow B against the previous species (arrow A_2) in Fig. 3 based on myth. The schema of each evolutionary theory is shown in Figs. 3, 4, 5 and 6. The horizontal axis represents time, and the vertical axis represents the different species. Because myth is ill-founded, the process that a new species *B* was born is not clear in it; therefore, myth is equivalent to part *K* (random state) in Fig. 2. However, it obeys chaos theory because humans considered it; thus, its position is not part *K* but the far-right

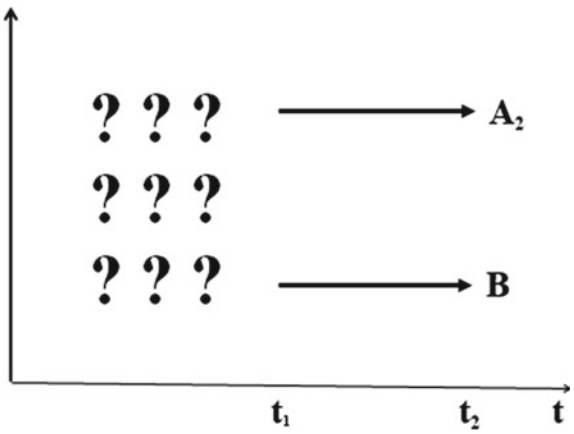


Fig. 3 Schema of mythic thinking. A new species is shown with arrow *B* against the previous species (arrow A_2)

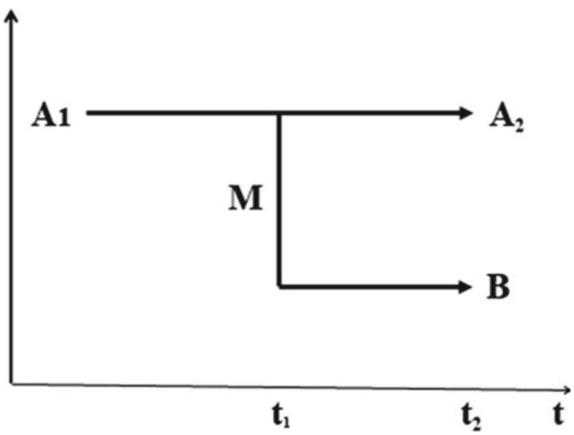


Fig. 4 Schema of assertive thinking. A new species separated from the old species (arrow A_1–A_2) is at the time t_1

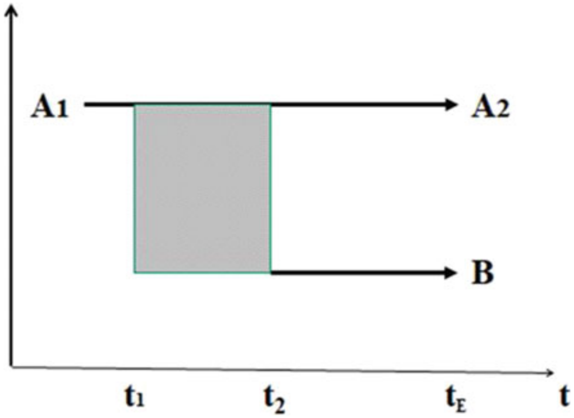


Fig. 5 Schema of non-assertive thinking. The non-assertive phenomena are shown as the rectangle box part on the right side of the time t_2

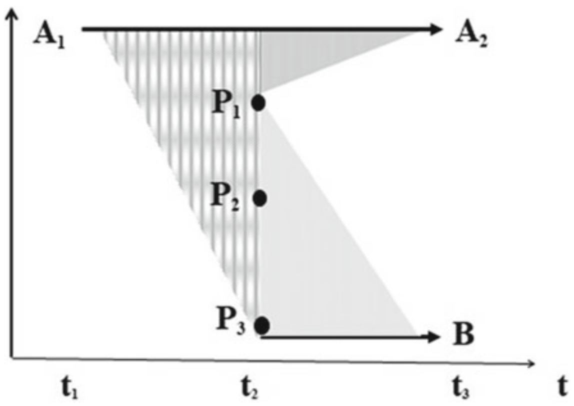


Fig. 6 Schema of decreasing entropy thinking. The birth of the new species is shown as a narrowing triangle

side of part *H*. This positioning is similar to that of myth in psychology, physics, and philosophy.

Creationism is similar to myth, and proof of evolution can be found in human embryos [16]. All humans repeat the same evolution process of 3.5 billion years by eight weeks after fertilization; the human heart cannot develop instantly with two atriums and two ventricles. As it is made by two atriums and one ventricle from one atrium and one ventricle, meaning that all humans evolved through the amphibian form from fish. If the repeating process stops at an incomplete state, a neonate with an anomaly or heart disease will be born. All humans experience the evolutionary process of fish, amphibians, reptiles, and mammals. Therefore, embryology denies that another entity has created each species.

7.2 The Relation Between Assertive Thinking and Darwin's Theory

Evidence separates Darwin's theory from myth [29]. Its schema is shown in Fig. 4. A new species separated from the old species (arrow A_1–A_2) is at the time t_1. It is considered that a new species B is born with mutation and therefore shown as a solid line M.

In the neutral theory of evolution [30], not one but many mutational changes are considered the center of evolution itself. Each species does never change overtime except in the moment of mutation. It means that there is no time of species except mutation; therefore, both theories are assertive (part E_b of Fig. 2). This positioning is similar to assertive thinking in psychology, physics, and philosophy.

7.3 The Relation Between Non-assertive Thinking and Imanishi's Theory

A schema of Imanishi's theory [31] is shown in Fig. 5. Imanishi stated, *"The species changes on the time that it must change."* His theory is non-assertive because it does not deny that species change with time and because it relates to some evidence. The central ability of evolution in living creatures is not mutation. Its non-assertive phenomena are shown as the rectangle box part on the right side of the time t_2. It is equivalent to the chaotic state in chaos theory, such as part G or H in Fig. 2. Because it obeys some evidence, its position is nearer the fixed state than the myth. This positioning is similar to non-assertive thinking in psychology, physics, and philosophy. However, the process with decreasing entropy that created a new species B is not clear.

7.4 The Relation Between Decreasing Entropy Thinking and Gene's Learning Evolution Theory

The thinking of living creatures relates to evolution because thinking is the same phenomenon with decreasing entropy as evolution, chaos theoretically. Your DNA will change with your exertions being decided by your thinking. It was proved that a father's experiences are transmitted to his children through sperm [32].

The blood pressure of adult giraffes is over 300 mmHg because of their long neck. Giraffes developed a complex pressure-regulation system in their upper neck, which only exists in their species, through the evolutionary process. This author thinks this system was hoped for by giraffes to protect their brains from a cerebral hemorrhage. In pregnancy, hypertensive information with medicine is transmitted to an embryo from the mother; this is equal to the heredity of diabetes mellitus with medicine. Receiving such medicine or not is the mother's will.

Thus, the child's DNA will change with the parent's experiences and thoughts. This fact can be explained by genetically transmitting adaptation information from one generation to the next. As for the explanation of embryology, the process of evolution is memorized in our genes and repeated in individual growth. True evolution is active and independent according to the will of the living creature [12, 16].

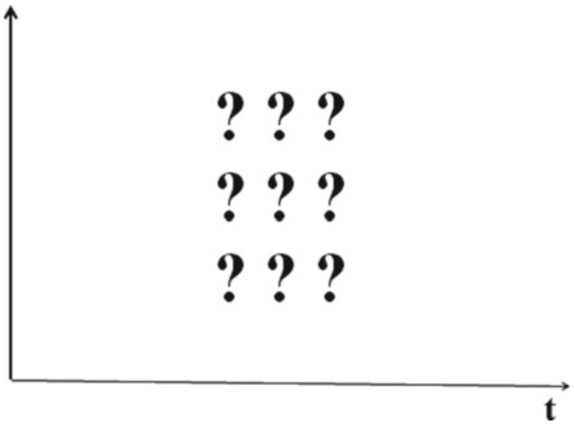
A schema of a new evolution theory is shown in Fig. 6. In gene's learning evolution theory, living creatures adapt to transformed environments and become new species with the gene's learning function [12, 13, 16]. This is shown as an expanding triangle, with the time that the old species adapt to a changing environment from t_1 until time t_2. The adaptive ability of Lamarckism [33] is a part of this theory. The new species' birth is shown as a narrowing triangle, with the time that the new species B is born with decreasing entropy [12, 13, 16, 34] from the time t_2 till the time t_3. This positioning is similar to decreasing entropy thinking in psychology, physics, and philosophy. The change representing decreasing entropy over the course of time can be explained with the arrow L in Fig. 2. Thus, the time of species clearly exists in this theory. Chaos theory can also demonstrate that a half-species does not exist; a species is equivalent to an incomplete fixed state such as the parts $E_{1–4}$ in Fig. 1. In chaos theory, these are fragmentary.

8 The Relation Between Chaos Theory and Development of Psychology

8.1 The Relation Between Psychology and Myth

Freud [35, 36] separated psychology from myth. Myth is shown as many question marks of Fig. 7 because its range and theoretical grounds are unclear. In Figs. 7, 8, 9, 10 and 11, the horizontal axis represents time and the vertical axis represents the different theories. Until Freud distinguished conscious and unconscious, psychology had not been considered an academic field.

Fig. 7 Schema of mythic thinking. Myth is shown as many question marks



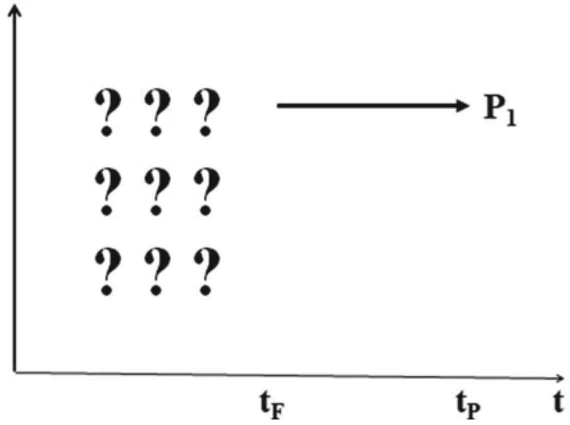


Fig. 8 Schema of assertive thinking. The theory of assertive thinking is shown as an arrow P_1

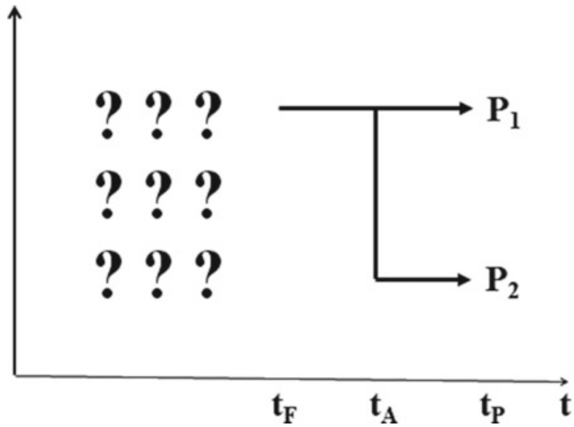


Fig. 9 Schema of assertive thinking. A new theory of assertive thinking is shown as a separate arrow P_2 from P_1 on time t_A

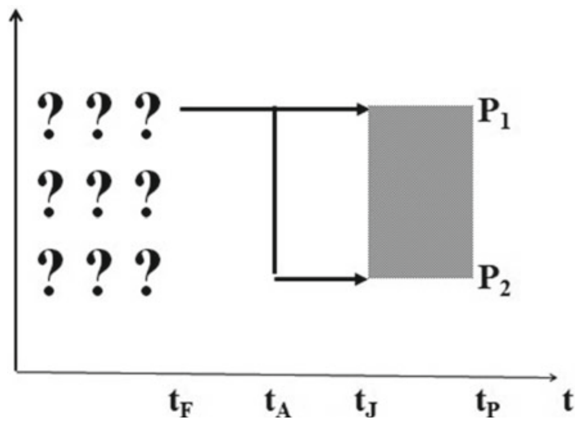


Fig. 10 Schema of non-assertive thinking. A theory of non-assertive thinking is shown as the rectangle box

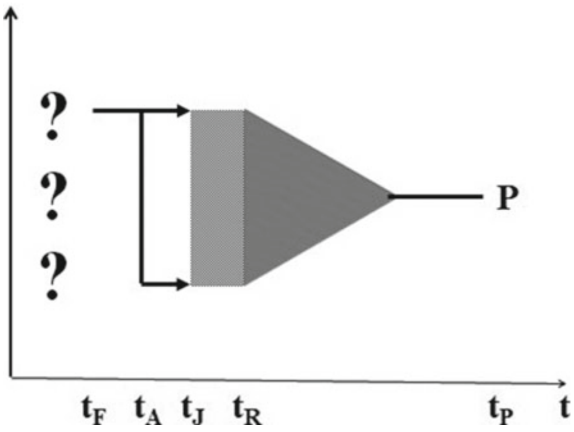


Fig. 11 Schema of decreasing entropy thinking. A theory of decreasing entropy thinking is shown as a triangle on the right side of time t_R

8.2 The Relation Between Assertive Thinking and Freud's Psychology

Because Freud [35, 36] first led psychology to become an academic field, his theory's schema is shown as an arrow P_1 in Fig. 8. He noticed the existence of unconsciousness on time t_F. However, he attributed all phenomena of human mentality to causes related to sex. He was not able to solve the problems of almost all patients because his theory is assertive.

8.3 The Relation Between Assertive Thinking and Adler's Psychology

Adler [37] considered human mentality to be the weak person's place; in other words, a person needs psychology because of their weak mind. His psychology was opposed to Freud's; however, its schema is shown as a separate arrow P_2 from P_1 on time t_A (Fig. 9) because it made its considerations based on Freud's psychology. However, his psychology could not provide an explanation for schizophrenia nor decreasing entropy.

8.4 The Relation Between Non-assertive Thinking and Jung's Psychology

Jung reported that schizophrenia could not be explained with assertive thinking [38, 39]. Assertive thinking is an incomplete fixed state, such as parts E_a and E_b in Fig. 2, while schizophrenia is similar to the proliferating chaotic thinking near the random state such as parts H or K in the same Figure [12]. Because Jung could never know chaos theory [17], he could not consider both fixed thinking and chaotic thinking. When the incomplete fixed (assertive) state cannot be considered, part G or H of Fig. 2 will be understood as non-assertive states with limited range. Therefore, Jung's psychology schema with some evidence is shown in Fig. 10 as the rectangle box. Jung noticed his theory on time t_{J}. Because it can only be explained with non-assertive thinking, Jung's theory was similar to quantum mechanics [40]; however, his theory could not explain all issues related to human mentality. It was reported that depression could be presented with the incomplete fixed state (parts E_a and E_b in Fig. 2) [12], and it can never be explained with non-assertive thinking—the human mind always changes. Finally, Jung's psychology did not explain decreasing entropy.

8.5 The Relation Between Decreasing Entropy Thinking and Rogers' Psychology

Rogers considered his counseling method [41]. It is a feature of his counseling that only the client knows the correct answer. Before Rogers' counseling, psychologists tried to correct the patients according to each theory. As a result, most of the patients' intentions were ignored. However, Rogers assisted in shifting the direction toward the clients, never controlling their minds. It was explained in Sect. 5.2 that human thinking can understand both incomplete fixed (parts E_a and E_b) and chaotic states (part G and H), as represented in Fig. 2; further, it is human development that allows rearranging thinking from chaotic to fixed states quickly. Rogers' counseling is equivalent to it; a schema of its process is shown as a triangle on the right side of time t_{R} in Fig. 11. The chaotic state of the client's thinking shifts to the fixed state with counseling.

Theoretically, entropy decreases in the client's thinking only through counseling. However, the client must have the ability to understand his/her problem and the hope to change his/her present condition. In other words, the effect of counseling is weak on patients with severe mental diseases because they cannot hope for their conditions to improve; therefore, they must not be treated only with counseling. Severe depression and schizophrenia are similar to the incomplete fixed state near the complete fixed state (part B) and the proliferating chaotic state near the random state (part K) in Fig. 2 [12]. Counseling must be used only with clients who can understand both incomplete fixed states and chaotic states; thus, it is better to implement counseling in the localized chaotic state (part G of Fig. 2). In short, counseling must be used to treat neurosis and so on.

9 The Relation Between Chaos Theory and Development of Physics

9.1 The Relation Between Physics and Myth

Until the emergence of Copernican theory [42], the Ptolemaic geocentric system theory [43] had been believed in medieval Europe de facto. The Ptolemaic theory was equivalent to a mythical theory based on poor observations; its schema is here equivalent to Fig. 7. In Figs. 7, 8, 9, 10 and 11 of physics, the horizontal axis represents time, and the vertical axis represents the different theories discussed.

9.2 The Relation Between Assertive Thinking and Newton's Equation

Both Copernicus and Newton [22] presented the Copernican theory and Newtonian gravitational equations based on many observations. In their equations, the energy change of a gravitational wave was not considered; thus, energy change over time does not exist. Therefore, all theories and equations are assertive and do not obey chaos theory because continuous covariation was not considered. A schema of their theories is shown in Fig. 8, where they are represented as the arrow P_1 and reported at the time t_F. Newton believed in the existence of absolute time that does not change with the environment. All phenomena could not be explained with his theories because Newton's opinion and subconscious psychodynamics [12] were mixed in with his equations. Thus, these were approximate equations describing a part of all phenomena within the totality of the cosmos.

9.3 The Relation Between Assertive Thinking and Einstein's Equation

Einstein [44] never considered Newton's equations nor many observations in the movement of heavenly bodies. As a result, he proposed a gravitational equation based on the constant speed of light. An observation from the opposite side of the movement was considered in his relative theory. The schema of his equation (arrow P_2) is different from that of Newton's equation (arrow P_1) in Fig. 9. In Einstein's theory, the energy change of an electromagnetic wave and of a gravitational wave could be considered. However, he denied this change, which made his theory assertive because the time of an electromagnetic wave and a gravitational wave does not exist. Einstein's equation thus presented a defective part [14, 15].

9.4 The Relation Between Non-assertive Thinking and Quantum Mechanics

It became clear that all phenomena with more detailed observations could not be explained with assertive theories such as those represented by Newton's and Einstein's equations. Quantum mechanics [45] was then born as a non-assertive theory. It is representative that the uncertainty principle was developed by Heisenberg [46] and that it relates an uncertain phenomenon to elementary particles. The observed state changes with observation and with the observation side. Because of this covariant relation, the uncertainty principle is part of chaos phenomena. However, the consideration given to "continuation" was insufficient. In quantum mechanics, all phenomena are presented by stochastic means. Because quantum mechanics relates to some evidence, it is closer to a localized chaotic state than the myth. Its schema of non-assertive phenomena [6, 10, 12, 13, 16] is shown as the rectangle box part on the right side of time t_J in Fig. 10. It is not random but localized because it is based on some evidence. However, quantum mechanics is an approximate theory; it does not explain the existence and direction of time [12].

9.5 The Relation Between Decreasing Entropy Thinking and New Gravitational Equation

In current cosmology, the Big Bang [3], dark matter [4], and dark energy theories [5] are believed to be correct. Because all-natural phenomena are amenable to chaos theory, each energy state of light (electromagnetic wave) and of a gravitational wave always changes in their environment. However, it is a precondition of the Big Bang theory that the energy state of light does not change; according to it, a cause of redshift is the Doppler's effect the moment that light was emitted from the heavenly body. Time cannot exist in light because its energy after being emitted never changes [12, 21]. It exists only when light is emitted and received. The energy state of light has a continuous covariant relation to its environment because it is related to gravitation [44]. Therefore, it must obey chaos theory. Because chaos theory can explain time [12], it denies the Big Bang theory, which ignores the time course of light (electromagnetic wave). Moreover, the energy state of a gravitational wave is similar to that of light.

There are six fatal contradictions in the Big Bang, dark matter, and dark energy theories.

1. Cyanobacteria of 3.5 billion years ago are now living [47]. A living creature always evolves due to some environmental changes. Therefore, cyanobacteria's existence means that the present inertia force is almost the same as 3.5 billion years ago. If the universe is expanding according to the Big Bang theory, inertia force must decrease greatly;
2. The Big Bang's time had been previously changed from 12.7 billion years ago to 13.7 billion years ago [12]. In 1995, this author reported that the Big Bang's

time would be pushed back according to observations of more distant heavenly bodies [48];

3. The existence of heavenly bodies 14.46 ± 0.8 billion light-years away from earth was discovered [49], and the existence of carbon or oxygen in these bodies 13.3 billion light-years away was confirmed [50]. In the Big Bang theory, it is explained that all elements except hydrogen and helium were made by a supernova explosion. Therefore, six processes were required before their heavenly bodies emitted light;

   First, hydrogen and helium, being scattered with the Big Bang, had gathered with their gravitation.
   Second, many supernovae were born.
   Third, their supernovae exploded.
   Fourth, carbon, oxygen, and so on were made with the supernova explosions and subsequently scattered.
   Fifth, they gathered with their gravitation.
   Sixth, the heavenly bodies (13.3 billion light-years away), having emitted light, were made.

   In the Big Bang theory, the time for the six processes took only 0.5 billion years.

4. If scientists measure the Hubble constant in five different ways, they will get five radically different values for it [51];
5. Galaxy size can become infinite according to the old gravitational equations. However, all galaxy sizes are limited, and limited galaxy size can never be explained with dark matter and old gravitational equations [52]. For this explanation, a sudden change of gravitation is required in the new gravitational equation [24].
6. The force of dark matter is attractive on a galaxy radius level; however, it is repulsive on a universal level. When the attractive force of dark matter in the Milky Way galaxy is observed in a very distant galaxy, it must be considered as a repulsive force of dark energy [15]. It means that dark matter must change to dark energy according to the position that it is observed from. This is not scientific.

This author reported that Hubble's law is a phenomenon according to Eq. 3 [12, 48] and that the relation between music, picture, and fluctuation (1/*f*) [53] can be explained with Eq. 3 [54]. From Eq. 3,

$$E(t) = E(0)e^{kt} \tag{4}$$

Here, $\acute{\alpha}$ is the Hubble's constant.

$$e^{k} = 1 - \alpha. \quad (-1 \ll k < 0) \tag{5}$$

From Eqs. 4 and 5,

$$E(t) = E(0)e^{kt} \approx E(0)(1 - \alpha t) \quad (6)$$

$$E(t) = E(0)(1 - \alpha t) \quad (7)$$

Equation 7 is equal to Hubble's equation that the Big Bang theory is explained with. Thus, the Big Bang theory can be explained with an approximate equation of Eq. 3. This author reported that dark matter, dark energy, and limited galaxy size can be explained with the ignored energy deduced from this equation [14, 52]. The entropy of total academic fields decreases because the Big Bang, dark matter, and dark energy theories are unified with the new equation.

Figure 11 shows the Big Bang, dark matter, and dark energy theories as different academic fields on time t_R. The unification of such theories is shown as the triangle on the right side of time t_R, and the new gravitational equation is shown as arrow P.

10 The Relation Between Chaos Theory and Development of Philosophy

10.1 The Relation Between Philosophy and Myth

Before Greek philosophy [55], theories and opinions with no argument and rationale were considered relevant to society's philosophy. Because myth is ill-founded, its schema is shown in Fig. 7. In Figs. 7, 8, 9, 10 and 11, with the development of philosophy, the horizontal axis represents time, and the vertical axis represents the different theories.

10.2 The Relation Between Assertive Thinking and Greek Philosophy

In Greek philosophy, words were defined, and theories were clarified with arguments and rationale. They are assertive because they do not change over time. Their schema is shown as the arrow P_1 in Fig. 8. Hilbert's formalism [56] is an assertive type, too, as formalism applies thinking and opinion to fixed expressions.

In Greek philosophy, dialectics obeying chaos theory were considered. In some cases of dialectics, entropy may surely decrease [12, 28]; however, dialectics demanded not correct solutions but moderation ("synthesis") [57]. If "antithesis" is almost automatically decided by "thesis," a relation between "thesis" and "antithesis" is not a covariation, and dialectic is not always continuous. Such cases do not obey chaos theory, and entropy does not decrease with dialectic if the parties share no common purpose.

Therefore, most philosophers now think that "dialectics is wrong" [12, 58–60].

10.3 The Relation Between Non-assertive Thinking and Russell's Philosophy

Non-assertive theories were considered by Russell [61], Brouwer [62], and so on. Russell's phenomenology and Brouwer's intuitionism recognized changes over time. However, the process of decreasing entropy in them is not clear. Because they relate to some basis, they become localized rather than random; therefore, a schema of their thinking is shown as the rectangle box part on the right side of the time t_J in Fig. 10. It is equivalent to quantum mechanics and Jung's psychology.

Phenomenology is almost explained with only a chaotic state. All phenomena always have some continuous covariant relation to the others. Here mathematical principles and historical facts are excluded from phenomena because they never change with time. Thus, all phenomena obey chaos theory, and most phenomena relating to time are equivalent to the chaotic state. The contents, such as law, are equivalent to an incomplete fixed state (parts E_a and E_b). However, they can change to a chaotic state (parts G and H) with environmental changes. They are different from the product description equivalent to a complete fixed state (part B) with the impossibility of change. Therefore, most phenomena within reality can be explained with/by the chaotic state.

Intuitionism is explained with only a chaotic state. Intuition is not fixed and greatly changes over time. However, it is equivalent to the proliferating chaotic state because its process is not clear.

10.4 The Relation Between Chaos Theoretical Thinking and Russell's Five Postulates

Russell's five postulates [63] are explained via chaos theory. This author will summarize the key elements of this as follows:

- *"The postulate of quasi-permanence"* can be explained with the localized chaotic state, such as part G in Fig. 2. All solutions are similar in the localized chaotic state.
- *"The postulate of separable causal lines"* can be explained with the incomplete fixed state (parts E_a and E_b) to the chaotic state (parts G and H) in Fig. 2. Each different solution can be unified to a fixed solution according to condition change.
- *"The postulate of spatial–temporal continuity in causal lines"* can be explained with the relation between incomplete fixed state and chaotic state. There are four solutions in part E (points E_1, E_2, E_3, and E_4) of Fig. 1. According to the change of variable p, they can change to a chaotic state, and the solutions in a chaotic state cannot be distinguished from each fixed solution.
- *"The postulate of the common causal origin of similar structures ranged about a center, or, more simply, the structural postulate"* can be explained with a relation between incomplete fixed state (parts E_a and E_b) and chaotic state (parts

G and H) in Fig. 2. The method of shifting a chaotic state to a fixed state in rearranging thinking is used multiple times [7–13].

- *"The postulate of analogy"* can be explained with the relation between the incomplete fixed state (part D) and the chaotic state (arrows G and H) in Fig. 1. In Eq. 1, two solutions of part D (points D_1 and D_2) cannot exist with only one side. In most chaos equations, pair solutions such as this can exist. The numbers of pair solutions in the arrow E or Eq. 8 are not two but four.

$$\frac{Y(n+1)}{Z(m+1)} = \frac{p[1 - Y(n)]Y(n)}{p[1 - Z(m)]Z(m)} \tag{8}$$

There is no part of two solutions in Eq. 8.

Thus, Russell's postulates are similar to the characteristics of chaos theory. However, he could not have known it because it was not recognized as a common theory in the 1960s [17]. Russell's postulates alone are insufficient to explain chaos theory. Therefore, his thinking cannot be positioned as chaos theory based theoretical philosophy.

10.5 The Relation Between Decreasing Entropy Thinking and Philosophy

This author does not know of a concrete example of philosophy with decreasing entropy according to chaos theory. A logical ground with decreasing entropy is required in a new philosophy that is amenable to chaos theory; thinking in most fields will be included in it. The continuous covariant relation is a necessary condition of chaos theory. As stated previously, the pattern of decreasing entropy with the process of dialectics is similar to chaos theory; however, the purpose of dialectics is moderate, and it is different from a common purpose of thinking in many fields. Chaos theoretically, dialectics is insufficient, and moderation is not always right.

On the other hand, Leibniz defined space and time [6, 12, 21, 24]. His definition that *"time course is born with change"* suggested that continuous covariant relations exist. Because time can become a common variable within reality, all phenomena with a time course obey chaos theory. However, a common purpose for decreasing entropy was not present in Leibniz's philosophy, either. This author thinks that the common purpose of philosophy must be equal to that of a living creature. The new philosophy schema is shown as the arrow P on the right side of a triangle heading in Fig. 11. Times t_R is now.

The new philosophy must include the thinking of most academic fields, and it will become a chaos theory based theoretical philosophy.

11 Results

The parts being equivalent to the left side of Figs. 3 and 7 are equivalent to the far-right side of part H (near part K) in Fig. 2. The right sides of Figs. 3 and 8 are equivalent to part E_a in Fig. 2. Figures 4 and 9 are equivalent to part E_b in Fig. 2. Figures 5 and 10 are equivalent to part G or H in Fig. 2. Finally, Figs. 6 and 11 are equivalent to parts E_a, E_b, G, and H in Fig. 2. In this way, each schema presenting the developments of physics, evolution theory, psychology, and philosophy can be shown with the schema of chaos theory. Therefore, each of their developments can be explained with chaos theory.

12 Discussion

Chaos theoretically, two modalities of thinking type exist, a fixed type and a chaotic type. Thus, two are the directions with a change in thinking. One is direction increasing entropy, which destroys living creatures (e.g., natural selection). Indeed, species are selected in nature; however, they can never be created by natural selection. The other is direction decreasing entropy, equivalent to evolution (gene's learning evolution theory) and a part of thinking.

In all academic fields, each entropy partially decreases. However, the entropy of academic fields is increasing as a whole because of specialization. Therefore, the common purpose of academia may be lost with each specialization. In each academic field, the entropy change of the whole academia must be considered. This is similar to a relation between a conductor and many players in an orchestra.

Literature, music, and pictures can be explained with chaos theory; thus, human thinking obeys chaos theory [12, 64]. Such a unifying style of thinking considering time may be equivalent to the new academic fields.

13 Conclusion

Each development of evolution, psychology, physics, and philosophy can be explained with decreasing entropy as understood by chaos theory. The entropy of total academia increases with the specialization in each academic field. Such a unifying style of thinking considering time may be equivalent to the new academic fields. Its phenomenon with decreasing entropy is equal to a common purpose of living creatures.

Core Messages

- The existence and a direction of time are explained with chaos theory that fixed state and chaotic state exist in.
- Each development of evolutionary theory, psychology, physics, and philosophy can be determined as myth, fixed (assertive) thinking, chaotic (non-assertive) thinking, and decreasing entropy thinking.
- Because the time of light (electromagnetic wave) and a gravitational wave cannot exist in the Big Bang, dark matter, and dark energy theories, a new gravitational equation considering the course of time is reported according to chaos theory.
- A unifying style of thinking considering time may be equivalent to the new academic field.

References

1. Teresa RC (1985) Chaos in a three-variable model of an excitable cell. Phys D Non-lin Phenom:233–242. https://doi.org/10.1016/0167-2789(85)90060-0
2. Decroly O, Goldbeter A (1987) From simple to complexoscillatory behaviour: analysis of bursting in a multiply regulated biochemical system. J Theor Biol 124(2):219–250
3. Narlikar J (1991) What if the Big Bang didn't happen? New Sci:48–51
4. Lewin JD, Smith PF (1984) Can cosmic neutrinos be detected with superconductors? Astrophys Lett 24(2):59–67
5. Peebles PJE, Ratra P (2003) The cosmological constant and dark energy. Rev Mod Phys 75:559–606
6. Yanagisawa H (1996) Contradiction and development of modern science—beyond chaos theory, Kirishobo Co. Tokyo, pp 115–123. (in Japanese)
7. Yanagisawa H (2012) Relation of the chaos equation to god perceived by Pascal, Nietzsche and Nightingale. SGJ 3(2):207–212
8. Yanagisawa H (2014) Relation of the chaos equation to the SEIQoL-DW (schedule for the evaluation of individual quality of life-direct weighting) method. MJCP. http://cab.unime.it/journals/index.php/MJCP/article/view/974/pdf_40. Accessed 31 May 2020
9. Yanagisawa H (2015) Discovering equations in relation to the counseling process. MJCP 3 (1). http://cab.unime.it/journals/index.php/MJCP/article/view/1044/pdf_49. Accessed 31 May 2020
10. Yanagisawa H (2017) Decreasing entropy in thoughts and evolution: main ability related to inside nature. MJCP 5(2). http://cab.unime.it/journals/index.php/MJCP/article/view/1498/pdf. Accessed 31 May 2020
11. Yanagisawa H (2018) Chaos theory applied to irresponsible use of 'reception and syn-esthesia' by Japanese Public Health Nurses. Ind J Nurs Educ Administr 4(2):26–32
12. Yanagisawa H (2019) Relations between human thinking and chaos theory. Scientific & Academic Publishing, USA
13. Yanagisawa H (2020) Relationship between chaos theory and history of philosophy. GSJ. http://www.gsjournal.net/ReseachPapers-Philosophy/Download/8119. Accessed 31 May 2020

14. Yanagisawa H (2011) Energy ignored by Einstein: dark matter is not required. GSJ. http://www.gsjournal.net/files/4277_yanagisawa.pdf. Accessed 31 May 2020
15. Yanagisawa H (2015) Distance to turning point between dark matter's attractive force and dark energy's repulsive force. GSJ. http://www.gsjournal.net/Science-Journals/Research%20Papers-Cosmology/Download/584. Accessed 31 May 2020
16. Yanagisawa H (1992) Relation of heart (mind) to gene; the evolutionary theory with gene's learning function. Kiri-shobou, Tokyo. (in Japanese)
17. Kellert HS (1993) In the wake of chaos: unpredictable order in dynamical systems (Chicago, University of Chicago Press, 1993), cited in Crayton Bedford, The Case of Chaos, in Mathematics Teacher Magazine, Apr 1998
18. McBride N (2005) Chaos theory as a model for interpreting information systems in organizations. Inf Syst J 15:233–254
19. Feigenbaum MJ (1978) Quantitative universality for a class of nonlinear transformations. J Stat Phys 19(1):25–52
20. Stanford Encyclopedia of Philosophy (2009) Information processing and thermodynamic entropy. http://plato.stanford.edu/entries/information-entropy/. Accessed 31 May 2020
21. Evangelidis B (2018) Space and time as relations: the theoretical approach of Leibniz. Philosophy 3(2):9. https://doi.org/10.3390/philosophies3020009Accessed31May2020
22. Newton I (1687) Philosophiae Naturalis Principia Mathematica.Jussu Societatis Regiæac Typis Joseph Streater, London
23. Stanford Encyclopedia of Philosophy (2004) Newton's views on space, time, and motions. http://plato.stanford.edu/entries/newton-stm/. Accessed 31 May 2020
24. Stanford Encyclopedia of Philosophy (2007) Leibniz's philosophy of physics. http://plato.stanford.edu/entries/leibniz-physics/. Accessed 31 May 2020
25. Yanagisawa H (2004) Space time defined by stress equation. EJTP 2:11–15
26. Chan H, Bledsoe WW (1965) A man-machine facial recogbitior some preliminary results. Panoramic Research Inc., Palo Ca
27. Kanade T (1973) Picture processing by computer complex and recognition of human face (Ph. D. thesis), Kyoto University
28. Gilgen AR (2000) Common prescriptions for psychology derived from dialectical materialism and chaos theory. Psychol Rep 86(2):482–492
29. Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London
30. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626
31. Imanishi K (1938) Annot Zool Jpn 17:23–36; Mem Coll Kyoto Imp Uni Ser B 16:1–35 (1941). (in English)
32. Shea JM, Ryan WS, Carone BR et al (2015) Genetic and epigenetic variation, but not diet, shape the sperm methylome. Dev Cell 35:750–758
33. Lamarck JB (1809) Philosophi Zoologique. Bailliere Paris
34. Styer DF (2008) Entropy and evolution. Am J Phys 76:1031–1033
35. Freud S (1915) "The unconscious", SE (Standard Edition), vol 14, pp 159–204
36. Freud S (1963) General psychological theory. Paperes on metapsychology. Collier Books, Macmillan Publishing Company, New York
37. Adler J (1997) Lying, deceiving, or falsely implicating. J Philos 94:435–452
38. Jung CG (1939) On the psychogenesis of schizophrenia. J Ment Sci 85(358):999–1011
39. Silverstein SM (2014) Jung's views on causes and treatments of schizophrenia in light of current trends in cognitive neuroscience and psychotherapy research II: psychological research and treatment. J Anal Psychol 59(2):263–283. https://doi.org/10.1111/1468-5922.12073
40. Jung CG, Pauli W (1955) The interpretation of nature and the psych, pantheon, New York. Translated by P. Silz. German original Naturerklarung und Psyche, Zurich: Rascher, 1952
41. Rogers CR (1949) The attitude and orientation of the counselor in client-centered therapy. J Couns Psychol 13(2):82–94

42. Encyclopedia Britannica. Copernican system. http://www.britannica.com/Copernican-system. Accessed 31 May 2020
43. Evening Star: The Ptolemaic Model. Retrieved from http://www.polaris.iastate.edu/EveningStar/Unit2/unit2_sub1.htm. Accessed 31 May 2020
44. Einstein A (1916) Die Grundlage der allgemeinen Relativtatstheorie. Ann Phy 49:762–822
45. Solov'ev EA (2011) Classical approach in atomic physics. Eur Phys J D 65(3):331–351
46. Heisenberg W (1927) Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. Z Phys 43(3–4):172–198. (in German)
47. Schirrmeister BE, Sanchez-Baracaldo P, Wacey D (2016) Cyanobacterial evolution during the Precambrian. Int J Astrobiol 15:187–204
48. Yanagisawa H (1995) Relation of cosmos to living creature; equation of energy and space-time. Kiri-shobou, Tokyo. (in Japanese)
49. David C (2019) How can a star be older than the universe? Space mysteries: if the universe is 13.8 billion years old, how can a star be more than billion years old? Space.com. https://www.space.com-how-can-a-star-be-older-than-the-universe
50. Matsuoka K, Nagao T, Maiolino R et al (2011) Chemical properties in the most distant radio galaxy. A & A 532(L10). https://doi.org/10.1051/0004-6361/201117641
51. Freedman WL, Feng LL (1999) Determination of the Hubble constant. PNAS 96:11063–11064
52. Yanagisawa H (2011) Limited galaxy sizes shown by new gravitation equation. EJSR 53 (1):87–92
53. Anderson CM, Holroyd T, Bressler SL et al (1993) 1/f-like spectra in cortical and subcortical brain structures: a possible marker of behavioral state-dependent self-organization. ATP Conf Proc 285:737–740
54. Yanagisawa H (2004) Three equations of fluctuation (1/f) to feeling, Weber-Fechner's Law and Hubble's Law deduced from stress equation (dE/dt=kE). EJTP 3:32–36
55. Greek Philosophy. Retrieved from http://www3.northern.edu/marmorsa/grphilosnotes440.htm . Accessed 31 May 2020
56. Stanford Encyclopedia of Philosophy (2003) Hilbert's Program. http://plato.stanford.edu/entries/hilbert-program/. Accessed 31 May 2020
57. Stanford Encyclopedia of Philosophy (2016) Hegel's Dialectics. http://plato.stanford.edu/entries/hegel-dialectics/. Accessed 31 May 2020
58. Adorno WT (1966) Negative dialectics. Bloomsbury, American heritage dictionary of the English Language, 4th Edn. Updated 2009. Houghton Mifflin Company
59. Marx K, Engels F (1845) The holy family or critique of critical criticism. Against Bruno Bauer and Company, Frankfurt am Main
60. Mocombe PC (2016) The ant-dialectical signification of Erzulie Danthor and Bois Caiman of the Haitian revolution. Afr Identities 14(4):332–347
61. Stanford Encyclopedia of Philosophy (1995) Bertrand Russell. http://plato.stanford.edu/entries/russell/. Accessed 31 May 2020
62. Stanford Encyclopedia of Philosophy (2003) Luitzen Egbertus Jan Brouwer. http://plato.stanford.edu/entries/brouwer/. Accessed 31 May 2020
63. Herger P (2003) The Cambridge companion to Bertrand Russell. In: Griffin N (ed). Cambridge University Press, p 310
64. Yanagisawa H (2019) Chaos theoretical explanation to operating time and space in literature: a writer shows a chaos state intentionally in a novel and gives illusory joy of entropy decreasing to many readers. Social Science Research Network. http://www.ssrn.com/sol3/paper.cfm?abstrctid=3408257. Accessed 31 May 2020

Hideaki Yanagisawa is an industrial physician of Gunma factory in Marelli Corporation. He worked as a doctor of internal medicine until 2001 and at the time had been observing connections between diseases and the mind. In 2002, he started working in a public health center. And he was appointed to his current position in 2019.

He aims to explain cosmology, religion, evolution, and the mind through a singular theory. The arrow of time can be explained with not the second law of thermodynamics but the continuous covariation of chaos theory. Therefore, some theories of physics must be corrected.

Managerial Problems and Cultural Heritage in Information Technology

3

Paolo Rocchi and Stefano Za

Underlying our approach to this subject is our conviction that "computer science" is not a science and that its significance has little to do with computers.

Abelson and Sussman

Summary

This paper aims to examine the nexus of phenomena surrounding culture, information technology, and learning. In fact, despite the widespread adoption of computer systems, IS/IT projects often fail or give unsatisfactory results. A significant proportion of these failures derive from managers' behavior rather than from objective obstacles. The inadequate preparation of managers is responsible for the majority of failing work plans; thus, we decided to look into the cultural background of managers working as both IT providers and IT clients. The paper examines the cultural heritage of managers since, for decades, writers have noted that the essential notions of computer science are not clearly established. We still observe the confusing aspects of computing theories and find out how the fragmentary and abstract concepts learned at school tend to govern the conduct of managers with broad and practical responsibilities. The cultural heritage does not meet the expectations of modern management, in fact

P. Rocchi (✉)
IBM Italia, Roma, Italy
e-mail: procchi@luiss.it

LUISS University, Roma, Italy

S. Za
University of Chieti-Pescara, Pescara, Italy
e-mail: stefano.za@unich.it

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_3

informatics is far from becoming an authentic science. As a result of these findings, we issue a renewed appeal to improve theoretical research on computer science foundations to set up a theoretical framework that is exhaustive, consistent, and in alignment with the practical recommendations of research teams and international institutes.

Increasing confusion in IT by time passing.
The code of this chapter is 01001001 01010100.

1 Introduction

The Widespread Adoption of Information Technology

After the Second World War, society underwent a progressive transformation from traditional manufacturing, which was introduced by the industrial revolution, to an economy with an increased emphasis on informational activities. The productivity and competitiveness of operators began to depend mainly on their capacity to generate, process, and efficiently apply knowledge-based information. Computer systems and networks form the backbone of the modern information society and are now indispensable in almost all contexts, from public health to transport, agriculture, engineering, etc. The world around us is increasingly controlled by algorithms that predict and react to our needs and permeate our daily lives, and yet the Digital Age has not reached its apex and is still evolving.

The knowledge and rational use of information technology (IT) create innovative solutions and promote actions that integrate different areas of knowledge. Computing concepts have infused virtually every area of academic study.

How and why are computers gaining entrance everywhere?

This question turns out to be non-trivial since the so-called information technology consists of the hardware and the software technologies while the former includes the analog and digital techniques.

The term "analog" means "something that bears an analogy to something else," and the analog technology handles information which basically imitates import or copies natural signals. Analog systems exploit a wide assortment of mechanisms (optical, chemical, mechanical, electrical, etc.) in order to manipulate pieces of information coming from different areas. It can be said that an analog device is an "ad hoc" device, in the sense that engineers have devised it for the precise purpose of processing a specific type of signal. On the other hand, digital systems comply with general principles and share the same essential structure [1], and in view of this uniformity, it is natural to wonder how a standard and rather a rigid scheme can provide people with so many different solutions.

The flexibility and pervasiveness of computers arise from two main sources.

Electronic machines manipulate bits and execute binary operations, where both bits and circuits are standard. They conform to general rules and criteria; thus, various functions can be integrated into a single system. For example, the typewriter, the telephone, the pendulum clock, the camera, and the mechanical calculator—all used in the past—were based on analog technology. Each relied on a different principle, and these principles are separate and distinct from one another. Manufacturers designed and produced each analog solution for a specific purpose or situation and for no other since disparate mechanisms cannot be integrated. In contrast, digital functions can be merged since they share identical criteria and are compatible. We are all familiar with the modern cellular device that encompasses a telephone, a clock, a digital camera, a calculator, etc. The wealth of services that can be offered by a single electronic product encourages its diffusion throughout society.

Secondly, experts can tailor each digital system so that the digital standard model is able to serve a variety of needs. Engineers typically apply the software (SW) technology which intervenes after the hardware (HW). In spite of the great emphasis which highlights the exterior differences and keeps HW apart from SW, the hardware and the software strongly make and take sense only united to one another.

We now examine the pragmatic methods that allow computers to be adapted to different environments.

(HW) Electronic engineers design two principal types of HW architecture:

- "Special systems' (SPs), which are equipped with special and unusual HW components, fulfill very particular duties. The large assortment of SPs in existence has arisen in response to the distinct needs of scientists, doctors, businessmen, drivers, technicians, etc. For example, computed tomography is used in hospitals; industrial robots operate in factories; embedded systems are adopted in cars; and global positioning systems are connected to satellites.
- "General purpose" systems (GPs) include standard components and consist of unified models that are sold by manufacturers on the market, e.g., mobile phones, tablets, and laptops. These circuits perform basic computing tasks.

(SW) There is little need of SW for a SP since it has been prepared to perform a specific task, instead GPs encompassing uniform HW elements, undergo a significant second phase of preparation. Software programmers install several packages in a GP system, and in this way, they adjust the system so that it can execute all the functions required by the customers. The amount of software programs means that GP systems, which are standard from a physical perspective, can become capable of running in countless areas of application. Moreover, each set of installed programs undergoes continuous maintenance to update the services provided.

Table 1 System architectures and methods of preparation

Digital technologies		
Hardware	Special systems	General purpose systems
Software	Modest software development	Strong software development

In summary, HW and SW technologies work in succession and create systems performing an infinite spectrum of functions. That is why digital solutions prove to be much more flexible than analog. If we rid ourselves of technical myths, we discover how HW and SW are used together to reach the same goal and create the electronic devices used in every corner of the world. The coupled manufacturing processes shown in Table 1 make the reader conscious of how the current invasion of computers takes its origin. They illustrate the inherent capability of IT to evolve and adapt to a wide range of situations.

2 Enduring Failures

The “IS/IT project” is defined as the collaborative enterprise that designs and implements the so-called information system with the support of information technology (IS/IT). The interdisciplinary organization factually installs computer systems and arranges all the involved elements (human operators, components, etc.) necessary for the planned tasks. For example, computed tomography involves hospital staff, patients, and the hospital ward, all of which make up an information system assisted by IT. In practice, IS/IT projects improve on an existing service or product or may introduce something completely new. E-commerce platforms and social networks are popular examples of IT solutions which have enhanced the performance, capability, and flexibility of companies, the daily life of individuals, institutions, etc.

IS/IT projects are responsible for bringing the modern digital society into being. They play a leading role, and it is surprising that most projects raise criticism and dissatisfaction for decades. Numerous articles, pamphlets, and administration documents have been written about the delusive performances of IS/IT projects registered worldwide. Surveys have been conducted in both advanced and underdeveloped countries, in both traditional and innovative sectors of activity, and all of them illustrate a rather uniform sad landscape.

The figures of the reports fluctuate, yet it may be said that approximately twenty percent of IS/IT projects fail, in that they are either canceled or not used (Fig. 1). About half of all projects involve challenges; for example, they are completed but are late or over budget or give unsatisfactory results. Only around thirty percent of projects meet all the relevant targets [2]. These numbers are anything but trivial and

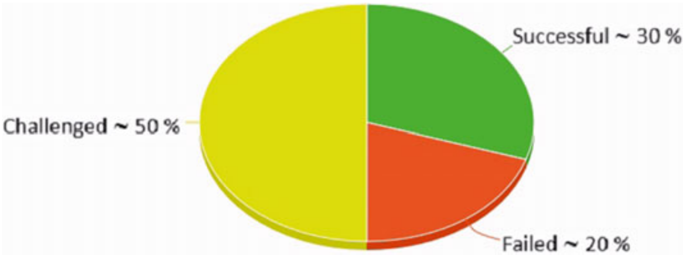


Fig. 1 Performance of IS/IT projects

mean that numerous organizations, both private and public, are wasting millions of dollars on failed projects [3]. These failures slow the progress of the digital society, which sometimes struggles to move forward.

Experts have explored the determinants of IS/IT project faults and have classified them according to various criteria.

The details change within the investigations and perspectives of researchers, nonetheless authors tend to converge toward three major factors of failure [5] (Fig. 2).

Process errors X: Most failure determinants are mistakes made by managers in relation to processes, e.g., insufficient planning, lack of resources, and changing requirements and specifications.

People errors Y: This group includes defects of managers in relation to people, e.g., unrealistic expectations, lack of user involvement, lack of IT management, and ineffective stakeholder management.

Mixed errors Z: The remaining factors derive from a variety of causes not belonging to the previous groups: technical, legal, operational, etc.

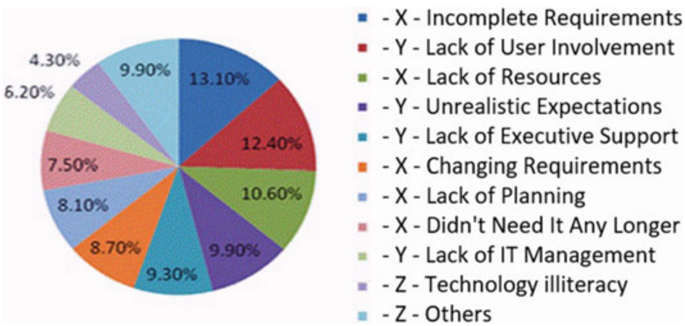


Fig. 2 Failure factors arranged in order of magnitude. *Source* Project Smart 2014; Figure of public domain

The summation of X factors in Fig. 2 approximately equals $\sim 48\%$; the summation of Y factors $\sim 38\%$ and Z factors reach $\sim 14\%$. The reader can note how most faults do not have objective and material origins. Most errors come from the ineffectual behavior of managers in relation to the processes, the stakeholders, and collaborators.

In this paper, the term “manager” is used to denote whoever is responsible for a given project, e.g. technical and non-technical staff, executives, project leaders, etc.

3 Efforts of Researchers and Practitioners

In the early beginnings, computer systems were used in universities and research centers. As soon as experts went beyond abstract computations and began to design applications for practical purposes, commentators observed the first “computer disasters”. The book of Glass [6] offers an earlier testimony of uncontrolled plans of work. He illustrates some heavy failures about the digitalization of society from a time long before computers had permeated our living, and many of the issues discussed by the author are still valid concerns today.

In the initial period of the digital era, the scientific community focused on the technical difficulties and proposed methodological solutions to the so-called software problem. They devised the methods for software development labeled as structured, waterfall, prototyping, object-oriented, and agile to mention a few of the most popular [7].

As time pass, scholars, and practitioners discovered that the spectrum of issues extended far beyond the purely technical perspective. They became aware of the multiple factors causing the project failures which include the ineffective usability of applications, the obscure language of communication, the ignored aims of the stakeholders on the part of managers, and so forth. Researchers started to move from technical-centered inquiries toward various directions in order to neutralize those negative factors. The authors have created immense literature which cannot be summarized here. We randomly mention writers who have examined the proactive and supportive human relations capable of improving user’s satisfaction. Others noticed users’ lack of preparation and how they do not significantly perceive the output quality. The hierarchical levels of management which establish the positions of the project members have been analyzed in different contexts. Some scholars provide help by means of a general theory. Professionals present the best practices and experienced case studies with the intention of offering illuminating examples to the leaders in the trouble. Experts of project management have taught how to improve the methods in use, how to optimize the modes of planning, organizing and delineating responsibility for the completion of the project goals, and how to refine organizational politics [8]. The International Organization for

Standardization (ISO) together with the International Electrotechnical Commission (IEC) delivered various publications with the scope of fixing the essential quality attributes of the software production and the metrics necessary to qualify them.

Numerous journals, conferences, symposia, seminars, etc., have addressed the various aspects of troubled IS/IT projects. Despite notable research efforts and economic investments made since the 1970s, the behavior of managers has improved only slightly, as the empirical data cited in the previous section show. Failing projects continue to lose billions of dollars worldwide and to create significant obstacles to the progress of IT in modern society. What is worst, researchers have not become sufficiently aware of the reasons for this global phenomenon. Different hypotheses have been suggested to explain the very slow progress in this field, but the scientific community did not reach a consensus.

4 Why Do Managers Fall Short of Expectations?

Troubled projects occur in every sector. It may be that a civil engineer designs an unstable bridge, technicians built up automobiles full of defects, and chemical processes pollute the neighboring environment. However, the managers do not systematically make mistakes like IT leaders. IS/IT runaway projects do not belong to a few teams, to a kind of organization, or a special institution. The inventoried problems occur everywhere in the world and repeat for decades. It is natural to wonder:

Why do managers fall short of expectations worldwide and over such a long period?

What have they equally?

We notice that IS/IT managers work in all five continents. They serve private companies and public bodies and install computers in various fields. They are not from a particular social class or age group and are both junior and senior professionals. They live in different habitats and come from different anthropological areas and levels. Consequently, we are inclined to conclude that these managers have nothing in common except for the computing culture in which they work.

Although they have not taken the same lessons in informatics, followed the same courses, or read the same books, they share the same conceptual basis in terms of computing. All IS/IT managers have an understanding of the same fundamental notions, including Turing machines, recursive theory, Shannon entropy and so forth, and managers who do not have technical expertise also share this cultural legacy. Lawyers, economists, and humanists who are responsible for projects have taken lessons on the basics of computer science (CS) out of necessity.

IT has rapidly developed over recent decades, and in consequence, IS/IT managers have learned different technical concepts. Solely the fundamentals of CS have not varied, and it seems reasonable to conclude that the ineffective behaviors of

IS/IT leaders result from the only aspect that they share: the basic notions of computing that they learned at school.

This judgment impeaches the bad preparation of managers; however, schools and universities do not invent the subject contents to expound. There is a certain cultural hierarchy: lessons derived from the achievements of academicians and researchers, and hence, the deficiencies in the scientific education of managers lead us to go through CS.

5 The Search for a Discipline

In the 1940s, "computing machines" were presented to customers as tools for mathematical operations, and it was natural to call computing the study of computer processes. Within a few years, it had become clear that these systems could be used for more than just calculations. When researchers succeeded in overcoming the original math-centric focus, they gained a more complex perspective and began to outline the basis of this new science.

In 1957, the German engineer Karl Steinbuch created the term ``Informatik' by compressing the words "information" and "automatische". The French followed with informatique, the English with informatics, the Slavic with informatika, the Italian and Dutch with informatica, and so forth. In the year 1959, the term computer science appeared in an article by Louis Fein in Communications of the ACM. Later, the scientific community put forward more sophisticated labels such as computing science, datalogy, computics, and the science of algorithms. All this seems a question about terminology, but we think not just that. Anyone contrives the word which better reflects his mind, and the missing uniform terminology echoes the critical intellectual state of the field. Recently, Tedre [9] has written a fine account of the long-term discussion about the essence of informatics and notices how computer experts wrestled with the core of CS throughout its entire history:

> There is no consensus on what computing as a discipline is actually comprised of. Asking ten computing researchers what computing as discipline is will yield ten different answers.

Abraham Kandel [10] was one of the first to complain about the cultural vacuum surrounding computers:

> It is quite obvious that there is no effective theory of computer science as such. In fact, there are no effective models of computers.

Knuth [11] expressed doubts in the structure and the ultimate purposes of CS:

> Having surveyed the relationships of computer science with other disciplines, it remains to answer the basic questions: What is the central core of the subject? What is it that distinguishes it from the separate subjects with which it is related? What is the linking thread which gathers these disparate branches into a single discipline?

Several writers share this criticism. They manifest dissatisfaction about the fragmentation of CS and the lack of a unified frame. They have written so many works as we confine ourselves to a few citations. Denning and Martell note how mature disciplines, e.g., mechanics and chemistry—are based on grand and unitive theories, whereas computing misses general principles. Henry Schaffer, professor emeritus at North Caroline State University, claims:

> [Computer science is] a conglomeration of many small areas that we pretend are a unified whole. Spreadsheets and cryptography are not as close as apples and oranges.

Hassan [12] pinpoints that what differentiates a discipline from a multidisciplinary field of interest is the development of a unique and consistent discourse, and Baldwin [13] blames the cultural discontinuity:

> Computing's fragmentation is well under way and is an unavoidable consequence of its maturation. However, fragmentation doesn't have to mean a collection of technology applications with no core science.

Hayes [14] warns the unscientific status of computer science:

> The resulting Balkanization of computing seems unwise and unhealthy, a recipe for reinventing wheels and making the same mistake three times over. Calls for unification go back at least 45 years, but the estrangement continues. (…) I find the standoff deeply frustrating.

Overall, several authors agree on the idea that the scientific community does not possess a common understanding of what constitutes informatics, its scope, and subject contents. The absence of meaningful and uniform concepts handicaps the progress of IT technology. Freeman and Hart [15] underscore how this cultural heritage does not provide adequate support to practitioners:

> Computer science and engineering needs an intellectually rigorous, analytical, teachable design process to ensure development of systems we all can live with. The core consideration for a software-intensive system is the software itself, and other approaches to systematizing design have yet to solve the 'software problem', which won't be solved until software design is understood scientifically.

Many researchers have highlighted the problems that arise due to a lack of the rigor which should characterize the study and development of SW, in addition, confusing notions also influence third parties such as customers, users, and stakeholders. Swanson and Fouad [16] note the consequences of the use of unspecific concepts:

> Without a theoretical framework to guide us, we would find it very difficult to make sense of the information clients might bring to us about their work-related problems. To return to the map or GPS analogy, we consult a road map before we leave on a trip to know the best way to get to our destination; without a map, we may wander aimlessly. So a good theory helps us represent reality, understand behavior, and assist clients in understanding their behavior.

All this matches with the root cause of the ineffective managers' behaviors that has been set in the previous sections.

6 The Source of Knowledge

The critical remarks quoted above make some serious points, and we therefore need to go more deeply into theoretical computer science (TCS), which has the aim of establishing the core of informatics. The cultural heritage of computing is based on constructions which form the conceptual basis for both schools and universities. TCS should provide an authentic and realistic understanding of computer systems and thus professional training. TCS was also conceived as a template for the activities and technical expertise that underpin the design of technical artifacts.

TCS looks like a mighty collection of constructs that do not leave out any technical content. TCS has so many theories as it is not easy to census them precisely [17]. The constructions on computing exhibit three characteristics that can be grasped even at a glance; more precisely, they appear to be narrow, self-referential, and abstract.

A theory is a coherent group of propositions that is formulated to explain a set of real facts or phenomena. Since a theory aims to define and predict events occurring in the real world, the broader it is, the more events it can explain and hence the more useful it is. In contrast, a narrow construct that deals with only a few objects is of little utility to academicians or practitioners. These essential requisites are well-known by computer theorists, yet they focus on narrow areas. Even the names of theories in CS indicate that each one relates to a limited field. For example, the five research streams of computability—computational biology, computational geometry, computational number theory and computational complexity—relate to five topics of limited breadth. Further examples include queuing theory, which focuses on files of elements, and relational algebra, which is used to model a relational database and define queries. In summary, each part of TCS refers to a particular argument that turns out to be somewhat trivial from a cultural viewpoint.

Scientists have frequently discovered partial results by chance and have taken advantage of the opportunity to build on these discoveries, finally identifying a complete and consistent domain of knowledge. In contrast, the components of TCS do not form a coherent whole, and some aspects are disjointed even when they involve the same topic. The five computational theories mentioned above are separate from one from the other. Another example involves coding, Shannon theory, and cryptology, which deal with discrete signals and should be integrated, but instead give disjointed results. The components of every SW program are instructions and data, but the theories of algorithms and data structures are not integrated. Due to this separation, informatics consists of several concepts that have no formal logical links. This gives rise to a rather chaotic cultural landscape, which the thinkers cited in the previous section have highlighted.

Turing assumed that an SW program was the solution to a mathematical problem, and it can be said that all computing constructs share this abstract view. Thinkers have even debated whether informatics should be organized as a branch of mathematics [9], although modern computing deals with human communication, economics, games, cybersecurity, education, etc., and very rarely solves abstract problems. Wikipedia [18] provides a list of free and open source SW programs that gives a full account of the types of applications created by modern SW developers and subdivides them into groups. Of 32 categories, only two, involving mathematics and statistics, are aligned with Turing's ideas. His approach was suitable in the early Digital Age but has proved to be obsolete in modern computing. SW programs address abstract problems only as an exception to the rule, and ethereal interpretations of SW technology do not prepare students for the professional reality that they will be confronted with.

Some authors claim that SW programs are mathematical, since they manipulate ones and zeros. This argument, which offers support for an abstract vision of SW technology, has no basis, since bits are not numbers. Electronics teach us that bits are precise and distinct signals that ensure the effectiveness of binary circuits and hence are anything but abstract elements. In the early stages of the development of CS, scholars chose the symbols one and zero to represent bits, but in principle, they could have chosen any other pair of different characters; in fact, some authors represent bits with the letters A and B.

7 Doubts

Managers and practitioners as well attended the lessons of computer science and apprehend identical concepts, yet the latter often create valuable products and do not show unprofessional conduct such as that of IS/IT managers.

Why is that IT specialists work effectively, while IT managers do not?

7.1 Low and High Levels of Responsibility

An IT specialist codifies and tests a program, he protects the desktop against hackers, he installs a package to monitor the server, etc. He fulfills a particular duty in a narrow area, even when dealing with cutting-edge techniques. As a consequence, the fragmentary concepts of computing do not impair the performance of IT experts, who have limited ranges of responsibility. For instance, when an expert needs to design a program that handles concurrent tasks, he finds support from "parallel computing", a construction belonging to TCS. The narrow cultural horizon of this theory not only hinder work in this area but is a factor contributing to effectiveness. TCS can provide excellent preparation for IT specialists; the partial and abstract theories do not generate adverse consequences.

The jobs of IT managers clearly mismatch with those of IT specialists (Table 2). A manager has multiple duties which span different areas of competence. For example, a manager may need to analyze technology, resource needs, and market demand; assess the feasibility of the project; examine different technical solutions; oversee coding, installation, and testing operations; prepare budgets, bids, and contracts; and confer with the client management and discuss project specifications.

Table 2 Cultural hierarchy according to the ranges of action

Cultural levels
Interdisciplinary ↓ Disciplinary ↓ Specialist ↓ Fractioned

In other words, an IS/IT manager needs an interdisciplinary culture to support a broad area of responsibility. He must have a complete vision of the information system in order to prepare and master all the component aspects. A manager must have a 360° view of this work, and it is clear that the abstract and divided notions furnished by TCS hinder this task.

There is a relation between the culture of a professional and his range of action (Table 2). An interdisciplinary approach addresses multiple questions, a particular discipline centers on a consistent group of topics, while fractioned notions are of little advantage. The partial constructions underpinning informatics are problematic from an intellectual viewpoint since they mismatch with the managerial range of action. TCS gives reason to the statement of Sect. 4 which relates the ineffective conduct of project leaders to the basis of knowledge they got when students.

7.2 Outsourcing

In the beginning of the computing era, in-house SW development teams were the most natural resources for IS/IT projects and corporate entities running SW implementations. Both private and public institutions took this view and set up digital services using their own human resources.

As time passed, however, SW houses began to offer IT services to clients. Companies and institutions hired experts external to their organizations (Fig. 3), and outsourced projects seemed like a "silver bullet" against the problems caused by in-house projects. Sometimes SW houses devised attractive, ready-to-use solutions and offered them on the market. Customers enjoyed excellent services in this way, as they no longer needed to be responsible for project management. Outsourcing was greeted with significant enthusiasm. This novel approach gave the impression that it could overcome all of the challenges related to IS/IT projects, but this false impression did not last long. As soon as the contracted services needed to be updated, improved, or required other interventions, the determinants of troubled projects again came to light, and the benefits of this outsourcing practice vanished.

The flawed cultural heritage analyzed above demonstrates that outsourcing does not address the root causes of troubled projects. All project leaders, whether internal or external to the organization, have the same weak points and defects. An outsourced service works perfectly until the management does not have to employ cross-cultural concepts, which are typical of managerial responsibilities.

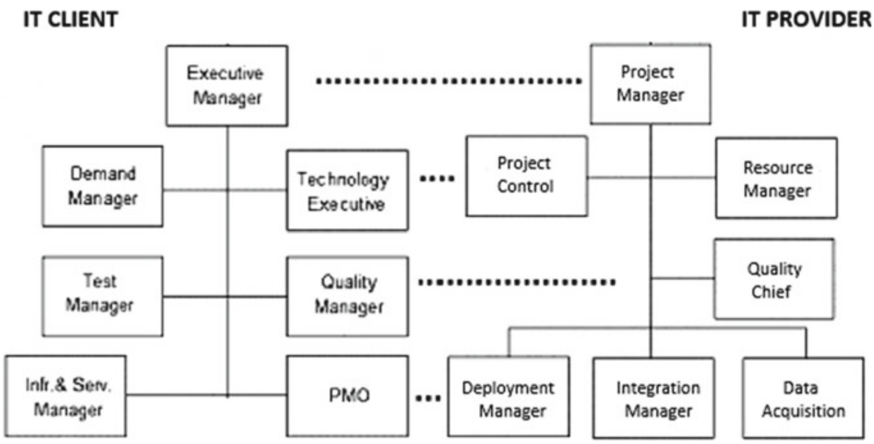


Fig. 3 Symmetrical levels of responsibilities in IT provider and IT client

7.3 Project Leaders and Stakeholders

A well-defined organization of the IT provider, whether internal or outsourced, can support the planned project. The members of the team are committed to the objectives of the work program at various levels; that is, they have a hierarchical ranking which dictates the relationships amongst the members of the IT provider and the IT client who requested the service. Good practice suggests that the structure of the provider should mirror that of the client at the upper levels [19]. The entities of the former should correspond to those of the latter, and this arrangement can help managers to cooperate in reaching common goals.

Cooperation requires transparent communication, simple expressions, and a shared vocabulary, although the fragmentary theories of TCS involve partial concepts and abstract definitions, which hamper clear communication at the managerial level. For example, participants may use different words to define a single object, and a single word may denote different elements. In addition to inadequate concepts, there are also incongruities in everyday jargon. The use of confusing language results in "human error", such as a lack of user involvement, unrealistic expectations of stakeholders, and a lack of IT management. The managers of the IT provider and those of the client both require clear terminology, so that they can communicate in an effective way. Weber [20] points out:

> (…) People in general have little grasp of what software is and does and have little or no awareness that there is such a thing as a 'software crisis'. Sensational exposé in the media (…) may indicate that awareness is increasing, even if in a distorted fashion.

In fact, TCS should generate shared notions and words; in turn, these encourage interaction of people; they keep discussion manageable and make easier to tackle problems. General and practical understanding should progressively correct the inappropriate behavior of managers working as IT providers and IT clients as well.

8 Interdisciplinary Notions

Various authors have advocated the definition of a theoretical frame that would give substance to CS as a unified discipline. Computing theorists should illuminate a broad area, rather than developing fragmentary constructions; they should integrate partial views rather than keep them apart from one other and should address practical questions rather than topics divorced from professional practice.

What kinds of conceptual references should be developed for managers?

A large part of TCS deals with the Turing machine (TM), and many studies have aimed to clarify, implement, develop, etc., the properties and performances of a TM. Despite this high level of attention, a TM has the scope of illustrating SW programming and focuses on technical particulars. TM disregards noteworthy features of computer systems such as the overall hardware structure, data storing, object programming, human–computer interaction, etc. IS/IT managers can be compared to navigators whose routes cross the oceans, while computing theorists are more like geographers who do not draw the map of the world but of a lake.

All the topics associated with IS/IT projects have yet to be placed under a single umbrella that enables both the managers of the provider and the client to dominate the multifold arguments that they are addressing. These leaders need to obtain a concise and complete knowledge of the artifacts for which they are responsible. They must use concepts that are as broad as their responsibilities (Table 3).

Experts from different fields have converged on the idea that CS deals with representation, implementation, manipulation, and communication of information. The literature recognizes to correlated elements:

(a) The concept of information,

(b) The concept of a system.

It can be said that the system (b) manufactures the product (a) when (b) can be mechanical, manual, or mixed. Both (a) and (b) are naturally linked to practical issues.

Although the terms information and system are used in everyday language and sound very intuitive and obvious, this is a somewhat false impression. Let us examine existing knowledge of these concepts.

(a) In the literature, there are hundreds of informal ideas about information; moreover, there are formal definitions that a recent bibliographical study has recently surveyed [21]. The study discovered 32 theories devised between 1900 and 2011 that have original traits and differ one from the other. The authors examined in this research not only disagree on the nature of the information but even on several essential aspects. One group conceives of information as something pertaining to the philosophical domain, while others focus on the properties of information

Table 3 Job responsibilities and intellectual contents

High level of responsibility → management tasks → control duties → interdisciplinary concepts
Low level of responsibility → specialist tasks → operational duties → disciplinary notions

stemming from technology, or a particular professional field. Around half of the contributions introduces mathematical definitions, while the remainder presented qualitative arguments. Shannon's theory, probably the best-known work, ignores semantics, which is a cardinal aspect of communication. Various authors made reference to the entropy function for measuring information, although this has different meanings and nuances within various frames. Finally, we recall the negationist theory of information devised by Maturana and Varela, two eminent biologists, who deny the existence of information as autonomous entity.

In addition, it is necessary to underscore, how semiotics presents the essential traits of informational elements. Semioticians furnish the basic notions to recognize whatever item of information, they call signifier and signified the base elements: The first is the material body of whatever sign which can be seen, heard, touched, etc.; the second is the mental or concrete entity denoted by the signifier. People use the nouns form and content as synonyms with the signifier and the signified in common conversation, this just to say that the semiotic concepts turn out to be very intuitive and conform with common sense. All of us often use the terms form and content without specific instructions. Electronics and SW experts also adopt semiotic notions in their daily work without being aware of them [22].

(b) In the 1950s, the biologist Ludwig von Bertalanffy outlined general system theory (GST) later furthered by writers coming from different areas. Thinkers such as William Ross Ashby, Margaret Mead, Gregory Bateson, and Charles West Churchman endeavored to draw a comprehensive frame of knowledge. They provided insights and considerations about the society, the human body, the organizations, and other complex entities of the modern world. GST begins from the idea that systems are inherently complicated, and they have an interdisciplinary nature which contrasts the usual scientific culture based on man-made and arbitrary divisions that set physics apart from chemistry, from biology, and so forth.

Systems thinking promised to write down a way to evolve toward an inclusive paradigm perfectly appropriate for managers. Thinkers have produced a wealth of philosophical works which, however, have had a negligible impact on the education of IT leaders so far.

The domain of systems has also a rigorous mathematical side whose origin dates to Henri Poincaré a pioneer researcher of dynamical systems. After the Second World War, this line of study gained increasing amounts of attention. The work of R. E. Kalman in the late fifties served as the starting point to the mathematical system theory [23]. Presently the literature goes through continuous and discrete systems, linear and nonlinear systems, stationery, complex, and stochastic systems. The roots of stochastic system inquiries intertwin with significant mathematical statements and models including the Poisson process, random walks, the Wiener process, the Lévy process besides the Markov processes. Active researchers look into the controllability, observability, and stability of systems and have arranged specialized constructs such as control theory, theory of automata, mathematical programming (or mathematical optimization), and ergodic systems. Computer experts have employed the mathematical system theory in image processing, signal processing, cryptography, telecommunications, and other areas.

In broad strokes, it may be said that system philosophers argue on great themes which, however, lie somewhat apart from the working practice. Mathematicians focus on professional problems but deal with narrow topics and do not enhance the preparation of IS/IT leaders. The two circles of systemic researchers devolve considerable intellectual energies, yet they did not much contribute to improve managers' understanding of computing so far.

9 Conclusion

People who use computers inherit a certain amount of scientific knowledge, and the present paper examines this cultural legacy.

In the introduction, we analyze the interplay of the hardware and software technologies that ensure the diffusion of computer systems in society. We then draw attention to the IS/IT projects that factually introduce computing in organizations and private living. Several surveys have revealed that a significant percentage of troubled projects are caused by a lack of professionalism in managers.

This study deduces a cause/effect relationship between the basic concepts that managers have acquired when young and their unsatisfactory professional conduct. In order to provide further evidence of this connection, we carefully analyze the subject contents of theoretical computer science and describe in detail how the limitations of TCS impact the management of IT providers and clients.

This work closes with a constructive appeal in favor of improvements in CS conceptualization, which will enhance the preparation of professionals in various fields. In the close it is worth mentioning the personal contribution to this plan of work [1, 24].

Core Messages

- Narrow and specialist concepts are insufficient when analyzing the global phenomena created by digital technology.
- Professionals and academicians need a top-down vision that is inherently interdisciplinary.
- The enduring failures of IS/IT projects raise cultural questions and suggest revisiting the current theoretical basis of computer science.
- Disparate and abstract theories can sometimes support specialists, but have been shown to be absolutely insufficient to support managerial responsibilities.
- IT providers and clients encounter communication difficulties, which are no more than cultural difficulties.
- The concepts of information and system should underpin the interdisciplinary culture of IS/IT managers.

- A definitive solution for IS/IT projects lies in the improvements in the cultural level of managers.

References

1. Rocchi P (2013) Logic of analog and digital machines, 2nd edn. Nova Science Publishers
2. Lynch J (2018) Project resolution benchmark report. Chaos reports, The Standish Group
3. Westland JC (2002) Systems complexity and the cost of errors in software development. J Syst Softw 62(1):1–9
4. Project Smart (2014) The Standish Group, Chaos Report. Available in June 2020 at https://www.projectsmart.co.uk/whitepapers/chaos-report.pdf
5. Nelson RR (2008) IT project management: infamous failures, classic mistakes, and best practices. MIS Q Exec 6(2):67–77
6. Glass RL (1997) Software runaways: monumental software disasters. Prentice Hall
7. Dobrica L, Niemela E (2002) A survey on software architecture analysis methods. IEEE Trans Softw Eng 28(7):638–653
8. Project Management Institute (PMI) (2013) A guide to the project management body of knowledge (PMBOK Guide), 5th edn. PMI
9. Tedre M (2015) The science of computing: shaping a discipline. CRC Press
10. Kandel A (1972) Computer science: a vicious circle. Commun ACM 15(6):470–471
11. Knuth D (1973) The art of computer programming, vol 1, 2nd edn. Addison Wesley Longman Publishing Co.
12. Hassan N (2006) Is information systems a discipline? A Foucauldian and Toulminian analysis. In: Proceedings of the 27th international conference on information systems, pp 425–439
13. Baldwin D (2011) Is computer science a relevant academic discipline for the 21st century? IEEE Comput 44(12):81–83
14. Hayes B (2015) Cultures of code. Am Sci 103(1):10–13
15. Freeman P, Hart D (2004) A science of design for software-intensive systems. Commun ACM 47(8):19–21
16. Swanson JL, Fouad AN (2015) Career theory and practice, 3rd edn. Sage
17. Hromkovic J (2011) Theoretical computer science, introduction to automata, computability, complexity, algorithmics, randomization, communication, and cryptography. Springer
18. Wikipedia (2019) List of free and open-source software packages. Available June 2020 at https://en.wikipedia.org/wiki/List_of_free_and_open-source_software_packages
19. Consoli C, Rocchi P, Spagnoletti P (2014) An empirical study of offshore software development: the case of a ticketing application. J Comput Inf Technol 22(4):267–275
20. Weber H (ed) (1997) The software factory challenge. IOS Press
21. Rocchi P, Resca A (2018) The creativity of authors in defining the concept of information. J Document 74(5):1074–1103
22. Rocchi P (2016) What information to measure? How to measure it? Kybernetes 45(5):1–16
23. Nijmeijer H, Schumacher JM (eds) (1989) Three decades of mathematical system theory: a collection of surveys at the occasion of the 50th birthday of Jan C. Willems. Lecture Notes in Control and Information Science, Springer
24. Rocchi P (2000) Technology + culture = software. IOS Press

Paolo Rocchi is a Docent Emeritus of IBM and Adjunct Professor at Luiss University (Roma, Italy). Present day, he is active in various fields including the Software Engineering, Theory of Probability, Quantum Mechanics, Education of Informatics, Information Theory, and Reliability Theory. Paolo has found intriguing results in the fields mentioned above and continues to explore these veins of research. Rocchi authored nearly one hundred and eighty works, including dozen books. More information is available at: http://www.edscuola.com/archivio/software/bit/eauthor.html

Stefano Za is Assistant Professor of Organization studies and Information Systems at "Gabriele d'Annunzio" University (Chieti-Pescara, Italy), adjunct professor at Luiss University, and visiting professor at the EM Strasbourg Business School (France). He is a personal member of ERCIS and Secretary of the Italian chapter of AIS (http://www.itais.org). Stefano has been a member of program committees and reviewer for several international conferences and journals in the domains of Information Systems and Organization Studies. His main research interest is the analysis and design of digital artifacts and organizational systems. He currently focuses on the digital innovations and business transformation affecting people and organizations in the digital ecosystem. Stefano was editor for several books and journal special issues. He has also published a book, papers on international conferences such as HICSS, ICIS, and ECIS, book series such as LNBIP and LNISO, and journals such as IEEE Wireless Communications, Personnel Review, GIQ, CAIS, I&M, JTAER, and BJET.

4 On the Essential Unity of Mathematics, Science, and Art: The Justice Evaluation Function and the Golden Number

Guillermina Jasso

It is a glorious feeling to perceive the unity of a complex of phenomena which appear as completely separate entities to direct sensory observation.

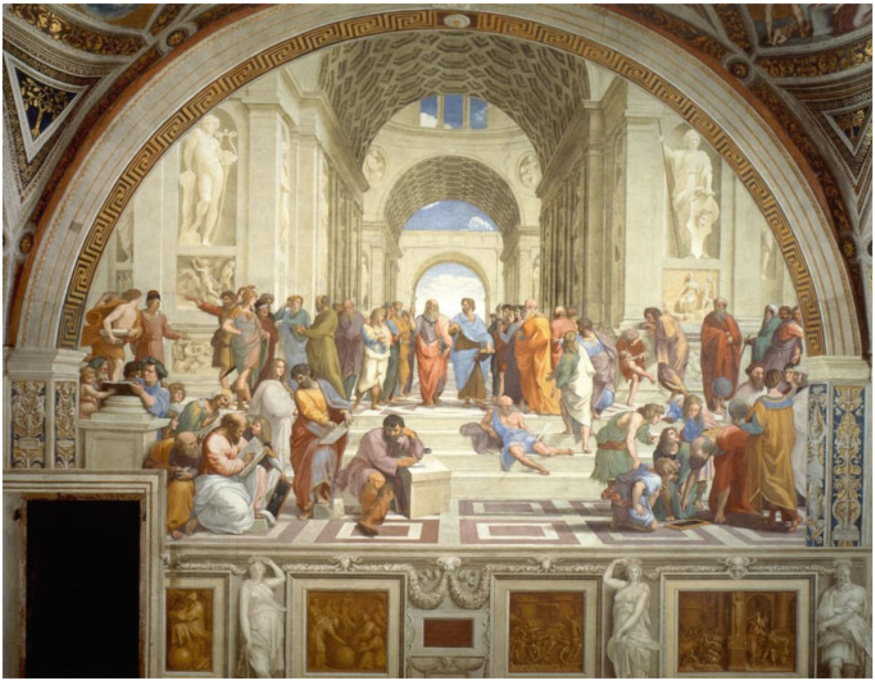
Albert Einstein (Quotation from a letter to Marcel Grossmann, written in Milan on 14 April 1901 [1].)

Summary

This paper describes how some basic scientific ideas—fairness/comparison/reference-dependence, deficiency and excess, loss and gain—coalesce, unexpectedly revealing a link to the Golden Number, which itself links mathematics and art. Fairness, comparison, and reference-dependence are mathematically equivalent. So are deficiency aversion—deficiency relative to the just amount is felt more keenly than comparable excess—and loss aversion—loss relative to the reference amount looms larger than comparable gain. Representing the outcomes (the justice evaluation J and the value V) by zero (for zero deficiency/excess or loss/gain), negative numbers (for deficiency or loss), and positive numbers (for excess or gain) leads naturally to a contrast between the negative and positive outcomes (deficiency and excess in one case, loss and gain in the other). This contrast can be expressed as a difference or a ratio. The ratio of the absolute value of the negative outcome to the positive outcome thus expresses the deficiency aversion coefficient or loss aversion coefficient. This ratio increases as the deficiency/excess or loss/gain k increases, crossing 2—deficiency is felt twice as keenly as excess and loss looms twice as large as gain—when k equals $\left(\sqrt{5}-1\right)/2$ (≈0.618) of the just or reference amount.

G. Jasso (✉)
Department of Sociology, New York University, 10012-9605 New York, US
e-mail: gj1@nyu.edu

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_4



School of Athens painted by Raphael in 1509–1511 in the Apostolic Palace in the Vatican. (Adapted with permission from https://upload.wikimedia.org/wikipedia/commons/4/49/%22The_School_of_Athens%22_by_Raffaello_Sanzio_da_Urbino.jpg).

The code of this chapter is 01000111 01101111 01101100 01100100 01100101 01101110 00100000 01001110 01110101 01101101 01100010 01100101 01110010.

1 Introduction

In the history of mathematics, science, and art, three words stand out: beauty, unity, surprise. The principles, discoveries, products, and other achievements radiate beauty. As understood at least since Pythagoras (c. 570–c. 495 BCE) and Plato (428/427 or 424/423–348/347 BCE), part of the beauty is the revealed unity. Moreover, these achievements come as a surprise, as do their beauty and unity. They may occur within or across domains, they may be large or small, monumental or modest, yet all delight by their beauty, unity, and surprise [2].[1]

[1] It is illuminating to remember that the subtitle of Thompson's [3] great mathematical classic *Calculus Made Easy* is *"Being a very-simplest introduction to those beautiful methods of reckoning which are generally called by the terrifying names of the differential calculus and the*

This chapter describes some small steps that end in a link across social science and mathematics. In the core story of this chapter, a simple question that arises both in the study of fairness and in the study of decisionmaking—When is deficiency relative to the just reward (or loss relative to the reference amount) felt twice as keenly as excess relative to the just reward (or gain relative to the reference amount)?—is unexpectedly answered by the Golden Number of mathematics. Specifically, deficiency (or loss) is felt twice as strongly as excess (or gain) when the deficiency (or loss) and the excess (or gain) are approximately 0.618 of the just reward (or reference amount).

This surprise signals a deep unifying element between the human sense of justice and mathematics. It signals also a deep bond between justice and beauty [13–18], a bond especially tight in languages where a single word signifies both justice and beauty—*fair*.

The chapter is organized as follows: Sect. 2 describes the justice evaluation and the justice evaluation function, summarizing justice theory and lingering on properties of the justice evaluation function, which will play a key part in the link to the Golden Number. Section 3 briefly reviews the history of the Golden Number and its universal esthetic appeal. Section 4 presents the justice approach to deficiency aversion and loss aversion, drawing out links between the justice evaluation function and the value function, mathematizing deficiency aversion and loss aversion, and representing the contrast between deficiency and excess and between loss and gain via both a difference and a ratio. Section 5 then analyzes the case in which the ratio of the absolute value of the negative outcome to the positive outcome equals 2—the case in which deficiency is felt twice as keenly as excess and loss is felt twice as strongly as gain—finding that this occurs when the deficiency/excess or loss/gain k equals $\left(\sqrt{5}-1\right)/2$ (≈ 0.618) of the just or reference amount. Section 6 discusses additional recent social science research which builds on the Golden Number. A short note concludes, followed by a short list of core messages.

2 The Justice Evaluation and the Justice Evaluation Function

The justice evaluation and the justice evaluation function are part of justice theory, a social science effort to understand the operation of the human sense of justice. Accordingly, we begin with a brief overview of justice theory.

---

integral calculus.” For brief social science overviews of beauty, unity, and surprise, see [4–10]. For examples from philosophy and literature, see [11, 12].

2.1 The Human Sense of Justice: Justice Theory

Around the clock and around the world, the sense of justice is at work. People form ideas about what is just, and they assess the justice or injustice of what they see around them. Both the ideas of justice and the assessments of injustice trigger a variety of individual and social processes, reaching virtually every area of the human experience. Justice concerns are pervasive, from the fierce children's cry of "It's not fair" to the melancholy reflection of heads of state that "Life is unfair" [19, 20]. It is therefore not surprising that understanding the sense of justice is a basic goal in social science, as pointed out in [21, p. 43–44], [22, p. xi–xii], and [23, 24].[2]

The framework for justice theory begins with four central questions, proceeds to identify three basic actors and four basic quantities, and finally embeds them in four basic processes represented by four basic functions. The four functions are deployed both theoretically and empirically. Theoretically, they play parts in both deductive and nondeductive theories, leading to testable predictions, including novel predictions, and testable propositions [28, 29]. Empirically, besides playing their obvious part in estimation and testing of the predictions and propositions, the four functions provide a foundation for measurement, for new data collection designs, for new data analysis protocols, and for interpretation of results.

2.1.1 Four Central Questions

There are four central questions in the study of justice:

i. What do individuals and societies think is just, and why?
ii. How do ideas of justice shape determination of actual situations?
iii. What is the magnitude of the injustice associated with given departures from perfect justice?
iv. What are the behavioral and social consequences of injustice?

Each question covers a family of questions, and each can be addressed both theoretically and empirically. The set of four questions, compiled by Jasso and Wegener [30, p. 398], integrates two earlier rival lists of three questions each [31, p. 1400] and [32, p. 155, 174].

2.1.2 Three Basic Actors

Three actors play fundamental parts in the sense of justice:

- Observer: The observer forms ideas of justice and judges the justice or injustice of specific actual situations;

[2] However, the sense of justice is not the only driver of behavior; as noted by Homans [23], status and power also play foundational parts. Thus, the sense of justice may not be universal—some people may be justice-oblivious, or may experience, or express, justice concerns only in some domains and not in others [25, 26]. Using Rayo and Becker's [27] evocative words, we may say that justice, status, and power are "carriers of happiness" and that the extent to which each occupies the mind and heart varies across people.

- Rewardee: The rewardee receives an amount of a good or bad being distributed or has a rank in the distribution of the good or bad; and
- Allocator: The allocator makes the distribution.

Some situations have only an observer and a rewardee. Sometimes one person plays all three parts or two of them. For example, when schoolchildren judge the fairness of the grades they receive in school, they are simultaneously observer and rewardee.

2.1.3 Four Basic Quantities

As is already evident from the four central questions, justice theory highlights four basic quantities:

i. Actual reward: The actual reward, denoted A, is the amount or level of the reward received by the rewardee;
ii. Just reward: The just reward, denoted C, is the amount or level of the reward the observer thinks just for the rewardee;
iii. Justice evaluation: The justice evaluation, denoted J, is the observer's assessment that the rewardee is justly or unjustly rewarded, and, if unjustly rewarded, whether underrewarded or overrewarded and to what degree; and
iv. Justice consequences: The observer's justice evaluation triggers many consequences, at both individual and social levels and touching vast areas of human experience.

Rewards may be cardinal, like earnings and wealth, or ordinal, like beauty and skills. Rewards of which more is preferred to less are called *goods*; rewards of which less is preferred to more are called *bads*. For example, for most observers, beauty and wealth are goods, and taxes and time in prison are bads. Cardinal goods are also called positive resources. Bads include both burdens (like chores) and punishments (the stuff of retributive justice). When the reward is cardinal, it is represented in the reward's own units (say, money or land or head of cattle); when the reward is ordinal, it is represented by relative ranks within a group or collectivity. People who value cardinal things are called *materialistic*; people who value ordinal things are called *nonmaterialistic*. Societies, too, are called materialistic and nonmaterialistic.[3]

The justice evaluation is represented by the full real-number line, with zero representing the point of perfect justice, negative numbers representing degrees of underreward, and positive numbers representing degrees of overreward.

[3] Elaborating footnote 2, if, as currently understood [25], status notices only the ordinal dimension of rewards, while justice and power notice both cardinal and ordinal dimensions, then there are five types of societies—justice-materialistic, justice-nonmaterialistic, status, power-materialistic, and power-nonmaterialistic—echoing Plato's insight. Thus, justice is active in two of the five types of societies.

2.1.4 Four Basic Functions

The actors and quantities are embedded in functions that address the basic questions:

i. Actual reward function: The allocator, guided by allocation rules, uses rewardee characteristics and other inputs to generate the actual reward for the rewardee;
ii. Just reward function: The observer, guided by justice principles, uses rewardee characteristics and other inputs to generate the just reward for the rewardee;
iii. Justice evaluation function: The observer compares the actual reward to the just reward, generating the justice evaluation; and
iv. Justice consequences function: The justice evaluation triggers a long train of justice consequences, possibly incorporating non-justice factors—stretching out to all domains of human behavior and the social life and giving distributive justice the character of a basic sociobehavioral force.

2.1.5 Remarks About Justice Theory

Before ending this brief overview of justice theory, it is useful to highlight several features.

First, note the parallel structure between the actual reward function and the just reward function. In the actual reward function, the allocator is guided by allocation rules to generate the actual reward, while in the just reward function, the observer is guided by principles of justice to generate the just reward;

Second, following Brickman et al. [33], the principles of justice include both principles of *microjustice*—pertaining to who should get what and why—and principles of *macrojustice*—pertaining to what the overall reward distribution should look like;

Third, the just reward and the justice evaluation are always observer-specific and rewardee-specific;

Fourth, observers demonstrate independence of mind, as enshrined in the fundamental principle owed independently to Hatfield [24, p. 152] and Friedman [34]: Justice is in the eye of the beholder;

Fifth, the actual reward, just reward, and justice evaluation lead not only to special functions (as above) but also to special distributions, such as the just reward distribution and the justice evaluation distribution, both of which can be succinctly summarized via observer-by-rewardee matrices [10, 35];

Sixth, though in this paper the rewardee is discussed for the case of an individual person, the justice apparatus scales up to the case in which the rewardee is a collectivity and the reward any of its characteristics, such as its resource endowment or its inequality. That is, the rewardee can be either a natural person or a corporate person;

Seventh, the justice evaluation function is tightly linked to two inequality measures:

- Atkinson's inequality (one minus the ratio of the geometric mean to the arithmetic mean); and
- Theil's mean logarithmic deviation (the log of the ratio of the arithmetic mean to the geometric mean);

Eighth, the justice evaluation function is used in the proof for a theorem stating that "*inequality in the distribution of a good is a bad, and inequality in the distribution of a bad is a good*";

Ninth, algebraic manipulation of the arithmetic mean of the justice evaluation distribution shows that overall injustice can be decomposed into injustice due to poverty and injustice due to inequality [35];

Tenth, justice theory has been generalized to the broader set of comparison processes, which includes not only relative deprivation, which may be thought of as fairness by another name, but also self-esteem, which pertains to the special case in which the rewardee is the same as the observer [36]; and

Lastly, justice/comparison theory yields a large number of testable predictions, including novel predictions, for example:

1. *Parents of two or more nontwin children will spend more of their toy budget at an annual giftgiving occasion than at the children's birthdays;*
2. *A thief's gain from theft is greater when stealing from a fellow group member than from an outsider, and this premium is greater in poor groups than in rich groups;*
3. *Veterans of wars fought away from home are more vulnerable to posttraumatic stress than veterans of wars fought on home soil;*
4. *Blind persons are less at risk of eating disorders than are sighted persons;*
5. *Inheritance tempers grief;*
6. *In groups where husbands earn more than their wives, marital cohesiveness increases with husbands' earnings inequality and wives' mean earnings and decreases with wives' earnings inequality and husbands' mean earnings; and*
7. *In materialistic societies with two warring subgroups, conflict severity increases as inequality increases.*

In this set of predictions, all were novel predictions at the time they were obtained except for Prediction 5, which exactly mirrors Cervantes' observation at the end of *Don Quixote*. As for empirical test, Prediction 1 is consistent with toy sales figures in the United States [37, p. 263], Prediction 3 with journalistic observations that Vietnamese veterans of the Vietnam War appear to be better adjusted than American veterans of the Vietnam War [38], and Prediction 6 with Bellou's [39]) finding that as male wage inequality increases, the divorce rate decreases. These and other predictions are amenable to a wide variety of tests, e.g., across cultures and historical periods. Thus, much is yet to be learned about the operation of the human sense of justice and of comparison processes more generally.

Figure 1 provides visualization of the world of distributive justice.

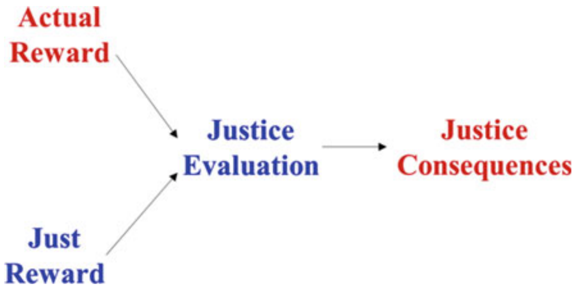


Fig. 1 The world of distributive justice. (Reproduced from Jasso et al. [29])

2.2 A Closer Look at the Justice Evaluation Function

The justice evaluation function (JEF) specifies the justice evaluation J as the logarithm of the ratio of the actual reward A to the just reward C (where "just" always means "just in the eyes of the observer"):

$$J = \theta \ln\left(\frac{A}{C}\right) \tag{1}$$

When the actual reward equals the just reward, the justice evaluation equals zero, the point of perfect justice. The parameter θ is called the *signature constant*. Its sign, called the *framing coefficient*, is positive for goods and negative for bads. Its absolute value, called the *expressiveness coefficient*, represents the observer's expressiveness. Things are framed as goods or bads by the observer. Formally, in the observer's eyes, a thing is a good if more is preferred to less, and a thing is a bad if less is preferred to more.

Thus, in the case of a good, when the actual reward is greater than the just reward, J is positive, indicating overreward, and when the actual reward is smaller than the just reward, J is negative, indicating underreward. Values of J close to zero indicate low degrees of injustice, and values far from zero indicate high degrees of injustice.

The hallmark of the justice evaluation function, embedded in Eq. (1), is that the justice evaluation depends on two variables, the actual reward and the just reward, and that the two have an opposite operation, such that, ignoring theta (or equivalently, for the case of a good), the first and second partial derivatives of J with respect to A are positive and negative, respectively, and the first and second partial derivatives of J with respect to C are negative and positive, respectively. Thus, in this case of a good, as A increases, holding C constant, the justice evaluation function is increasing and concave down, while as C increases, holding A constant, it is decreasing and concave up.[4]

The justice evaluation function has useful properties:

1. *exact mapping from combinations of A and C to J;*
2. *the outcome it yields is in justice units (not reward units);*
3. *integration of rival conceptions of J as a ratio and as a difference;*
4. *deficiency aversion,* viz., *deficiency is felt more keenly than comparable excess (and loss aversion,* viz., *losses are felt more keenly than gains);*
5. *scale invariance;*
6. *additivity, such that the effect of A on J is independent of the level of C, and conversely;*
7. *symmetry, such that interchanging A and C changes only the sign of J;*
8. *the log-ratio form of the justice evaluation function is the limiting form of the difference between two power functions, which both strengthens integration of the ratio and difference views and also integrates power-function and logarithmic approaches; and*
9. *a link between the justice evaluation function and the Golden Number, approximately 0.618, such that the loss aversion ratio equals 2 when the actual reward equals the just reward plus or minus the product of the just reward and the Golden Number.*

The first four properties were described in the original research report in 1978 [31]. Properties 5 and 6 were established in 1990 [36], properties 7 and 8 in 1996 [43]. Property 9 was first mentioned in 2006 [44] and more fully discussed in 2015 [45], but not comprehensively analyzed until this chapter.

The framing coefficient is always critically important, but the expressiveness coefficient, which is critically important in empirical analysis, can safely be ignored in purely theoretical analysis. Much of the work in this chapter is purely theoretical, and for simplicity, focuses on goods. Accordingly, the signature constant can be safely set to one. Only much later in the chapter will it be necessary to examine bads and as well to invoke the distinction between the *experienced* and the *expressed* justice evaluation [28, 30].

In the generalization to comparison processes, the actual reward is sometimes called the actual holding, the just reward the comparison holding (or more commonly the reference point), and the justice evaluation the comparison outcome and denoted Z. The comparison holding can arise from myriad quantities, such as self or

[4] The log-ratio specification of the justice evaluation function has rich intellectual roots, going back to Bernoulli's [40] utility function and Fechner's [41, 42] sensation function. However, the justice evaluation function differs in at least two important ways from its predecessors. Recall that in Bernoulli's function, utility is a logarithmic function of wealth, and in Fechner's function, subjective magnitude is a logarithmic function of the ratio of a physical stimulus to the (constant) lower absolute threshold for that stimulus. Thus, both Bernoulli's and Fechner's functions are functions of one variable, and they are concave functions. In contrast, the justice evaluation function is a function of two variables, and, as can be shown by inspection of the principal minors of its Hessian second derivative matrix, it is neither concave nor convex.

other's actual reward or function thereof, something in the past or envisioned or desired, a parameter of the actual reward distribution, and so on [46]. For example, in the foundational accounts owed to Marx and James, respectively, satisfaction arises from comparing the hut to the palace [47, pp. 84–85], and self-esteem arises from comparing success to pretensions [48, p. 200].

For comprehensiveness and clarity, the justice evaluation function and the comparison function are sometimes displayed in a fourfold classification, with both the general expression and the specific log-ratio specification for both the global version covering both goods and bads in a single expression and the conditional version with two branches, for goods and bads, respectively [35, p. 139]. Thus, for example, one of the four cells (Cell B.2) contains Eq. (1), the log-ratio form for both goods and bads. When only goods or only bads are of interest, the usual expressions are either adapted from the global expressions or use only one branch from the two-branch expression, as in Falk and Knell's [49, p. 418] and Jasso's [36, p. 380] general form for utility and comparison, respectively, in the case of a good, which also forms part of Cell A.1 in [35, p. 139].

Figure 2 displays graphs of the justice evaluation function, separately for goods and bads, and showing for each the graphs of J on the $\ln(A)$ and $\ln(C)$ components. As shown, the graphs of the two components are reflections of each other about the x-axis. Moreover, the graphs of the $\ln(A)$ components for a good and a bad are reflections of each other about the x-axis, as are the graphs of the $\ln(C)$ components for a good and a bad.

Note that when the just reward C is held constant, the justice evaluation function is fully depicted by the graph of $\ln(A)$, and this graph crosses the x-axis at the magnitude of A that equals C.[5]

The individual's time series of justice evaluations, called the justice profile, provides a picture of the interior justice life, with variation over a unit of time in the number and duration of distinct justice evaluations, their range, the jumps and dips between them, the gaps when the sense of justice is asleep, the means of the underrewarded and overrewarded truncates, etc. New questions arise concerning periodicity and links to specific rewards, to age and experience, and to the larger happiness profile insightfully discussed by Layard [25, p. 417, 50, pp. 367–370, 51].

3 The Golden Number

One idea permeates the Golden Number: beauty. There is beauty in proportions, there is beauty in numbers, and there is unrivaled beauty in the special number that represents a certain special proportion. To see this special proportion, consider a

[5] As shown, in the case of a good, the justice evaluation increases at a decreasing rate with the actual reward. Elaborating further on footnotes 2 and 3, the distinguishing feature of the three foundational engines of behavior is their rate of change, such that, continuing with the case of a good, as the actual reward increases, status increases at an increasing rate and power at a constant rate [25].

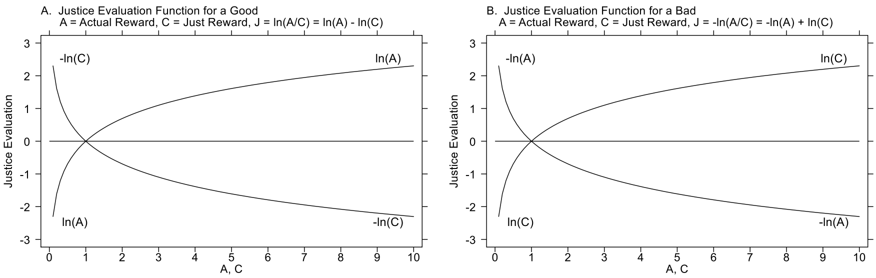


Fig. 2 Justice evaluation function for goods and bads

line AB. Suppose it is divided at an interior point P such that AP/PB equals PB/AB —that is, the ratio of the smaller to the larger line segment equals the ratio of the larger line segment to the whole line.

Ancient philosophers speculated that this relation among a quantity and two components of the quantity represented ideal proportions. As Plato [52, p. 448] writes in the *Timaeus*: "... the fairest bond is that which makes the most complete fusion of itself and the things which it combines; and proportion is best adapted to effect such a union. For whenever in any three numbers,... there is a mean, which is to the last term what the first term is to it; and again, when the mean is to the first term as the last term is to the mean—then the mean becoming first and last and the first and last both becoming means, they will all of them of necessity come to be the same, and having become the same with one another will be all one".

And mathematicians since the ancients used their tools to approximate the requisite point P for achieving that ideal set of proportions [13, 53, 54]. The first formal definition appears in Euclid's [55, p. 99] *Elements*, Book VI: "A straight line is said to have been *cut in extreme and mean ratio* when, as the whole line is to the greater segment, so is the greater to the less."

The solution for P turns out to be the positive solution of a breathtakingly simple and beautiful quadratic equation, viz., $x^2 + x = 1$ [15–17, 56, p. 104]:

$$\frac{\sqrt{5}-1}{2} \approx .618 \tag{2}$$

termed the Golden Number. Thus, if AB has a length of 1, the line is cut at 0.382, and the ratio of the smaller to the larger segment (0.382/0.618) equals 0.618, as does the ratio of the larger segment to the whole (0.618/1).

The history of the Golden Number has rich highlights. For example, it has acquired a portfolio of names, including Golden Ratio, Golden Section, Golden Mean (not to be confused with Aristotle's idea of the golden mean, in which moral virtue is a position between two extremes), Division in Extreme and Mean Ratio (DEMR), and Divine Proportion. Similarly, both the Golden Number and its reciprocal (approximately equal to 1.618) are widely used. At least since Plato, the relation is seen to be like a palindrome—it works the same whether we go from left to right or from right to left. It attracted the great astronomer Kepler, and its first decimal approximation seems to have been made by Michael Maestlin, Kepler's teacher. Reading its history is like reading the history of mathematics, the amazing growth of knowledge laid out before our eyes, a parade of geometric figures, the new Hindu-Arab numerals, algebra, sequences, and so on. Importantly, the Golden Number touches virtually every area of the human and physical worlds, where it is both found in nature and used purposefully in art, science, and engineering, among other endeavors [57, 58].

The Golden Number enthralls by its beauty and, echoing Plato, unifies everything it touches.

4 The Justice Approach to Deficiency Aversion and Loss Aversion

This section uses the justice approach to study two distinct situations: (i) the situation involving deficiency or excess relative to the just reward; and (ii) the situation involving loss or gain in the actual reward from Time 1 to Time 2.

4.1 Deficiency and Excess in Actual Reward, Relative to Just Reward

As before, let A denote the actual reward and C the just reward. Now let the actual reward equal the just reward plus or minus a constant k,

$$A = C_0 \pm k \tag{3}$$

where k is positive and less than C_0 ($0 < k < C_0$). The constant k is called the *deficiency* or *excess*.

It has been known since the first report of the justice evaluation function that it possesses the property that deficiency is felt more keenly than comparable excess [20], a property we may call *deficiency aversion*. Indeed, this is the property most closely associated with the JEF and the one that first attracted notice. The idea that underreward is felt more strongly than overreward goes back at least to Adams [59, p. 426], 15 years before the JEF was introduced. As Wagner and Berger [10, p. 719] observe, this was "a phenomenon all justice theorists assumed occurred, but one that had not been incorporated in their theories".

Formally, the deficiency aversion property is stated:

$$\left|J^{\text{Deficiency}}\right| > J^{\text{Excess}} \tag{4}$$

Its mathematical statement is written:

$$\left|\ln\left(\frac{C_0-k}{C_0}\right)\right| > \ln\left(\frac{C_0+k}{C_0}\right) \tag{5}$$

The discrepancy in J between deficiency and excess can be expressed in two ways, as a difference and as a ratio.

Define the difference D^J as the absolute value of the justice evaluation J in the deficiency case minus J in the excess case:

$$D^J = \left|\ln\left(\frac{C_0-k}{C_0}\right)\right| - \ln\left(\frac{C_0+k}{C_0}\right) \tag{6}$$

Upon algebraic manipulation, the difference D^J becomes:

$$D^J = \ln\left(C_0^2\right) - \ln\left(C_0^2 - k^2\right) \tag{7}$$

The difference D^J is always a positive quantity (as visible in the inequality in Eq. (4)), and it ranges to infinity.

Similarly, define the ratio R^J, which can also be called the deficiency aversion coefficient, as the ratio of abs (J) in the deficiency case to J in the excess case:

$$R^J = \frac{\left|\ln\left(\frac{C_0-k}{C_0}\right)\right|}{\ln\left(\frac{C_0+k}{C_0}\right)} \tag{8}$$

By algebraic manipulation, the ratio R^J becomes:

$$R^J = \frac{\ln\left(\frac{C_0}{C_0-k}\right)}{\ln\left(\frac{C_0+k}{C_0}\right)} \tag{9}$$

The ratio ranges from one to infinity.

Many special cases of both the difference D^J and the ratio R^J can be fruitfully examined. Below we focus on one particular question about the ratio R^J.

4.2 Loss and Gain in Actual Reward, Relative to Time 1 Actual Reward

A second situation arises when the actual reward A changes from Time 1 to Time 2, producing a change in the justice evaluation J from Time 1 to Time 2, denoted CJ. Denote the actual reward at Time 1 by A_0, let the just reward C remain constant from Time 1 to Time 2, and let the actual reward at Time 2 equal the Time 1 actual reward plus or minus a constant k:

$$A = A_0 \pm k \tag{10}$$

where k is positive and less than A_0 ($0 < k < A_0$). The constant k is called the *loss* or *gain*.

In this situation, we obtain the result, parallel to Eq. (4), that loss is felt more keenly than gain. This property is called *loss aversion,* and it was introduced by Kahneman and Tversky [60, 61], who observed that "losses loom larger than gains" [60, p. 279].

Formally, the loss aversion property is stated:

$$\left|CJ^{Loss}\right| > CJ^{Gain} \tag{11}$$

In the justice approach, its mathematical statement is written:

$$\left|\ln\left(\frac{A_0 - k}{A_0}\right)\right| > \ln\left(\frac{A_0 + k}{A_0}\right) \tag{12}$$

As with the discrepancy in J between deficiency and excess, the discrepancy in J between loss and gain can be expressed in two ways, as a difference and as a ratio.

Define the difference D^{CJ}:

$$D^{CJ} = \left|\ln\left(\frac{A_0 - k}{A_0}\right)\right| - \ln\left(\frac{A_0 + k}{A_0}\right) \tag{13}$$

Upon algebraic manipulation, the difference D^{CJ} becomes:

$$D^{CJ} = \ln\left(A_0^2\right) - \ln\left(A_0^2 - k^2\right) \tag{14}$$

The difference D^{CJ} is always a positive quantity (as visible in the inequality in (5.11), and it ranges to infinity.

Similarly, define the ratio R^{CJ}, which can also be called the loss aversion coefficient, as the ratio of abs (J) in the loss case to J in the gain case:

$$R^{CJ} = \frac{\left|\ln\left(\frac{A_0 - k}{A_0}\right)\right|}{\ln\left(\frac{A_0 + k}{A_0}\right)} \tag{15}$$

By algebraic manipulation, the ratio R^{CJ} becomes:

$$R^{CJ} = \frac{\ln\left(\frac{A_0}{A_0 - k}\right)}{\ln\left(\frac{A_0 + k}{A_0}\right)} \tag{16}$$

The ratio R^{CJ} ranges from one to infinity.

4.3 Summary of Justice Approach to Deficiency Aversion and Loss Aversion

The justice approach yields predictions for two distinct situations: (i) the situation involving deficiency or excess relative to the just reward; and (ii) the situation involving loss or gain in the actual reward from Time 1 to Time 2. The mathematical outcomes are identical. Justice theory predicts both that (i) deficiency is felt more keenly than comparable excess and (ii) loss is felt more keenly than comparable gain.

Moreover, the justice approach predicts the exact magnitudes by which deficiency (or loss) is felt more keenly than excess (or gain), doing so for both a difference representation of the discrepancy and a ratio representation of the discrepancy.

Of course, the interpretation is context-specific. The deficiency/excess results pertain to assessments of the actual reward relative to the just reward at a point in time. The loss/gain results pertain to assessments of the actual reward as it changes between two points in time.

For ease in contrasting results, using them, and building on them, Table 1 collects the main terms in the justice approach to deficiency aversion and loss aversion.

The sections below examine the case where the magnitudes of the ratios R^J and R^{CJ} equal 2. As will be seen, that is when the Golden Number appears.

5 Deficiency Aversion, Loss Aversion, and the Golden Number

5.1 Deficiency Aversion and the Golden Number

We turn now to examine the case in which deficiency is felt twice as keenly as comparable excess. This case is important because of empirical evidence that loss is felt twice as keenly as comparable gain [62, p. 1288, 63, pp. 1053–1054].

To analyze this case, we set the ratio R^J equal to 2 and solve for k. At the first step, we write:

$$\frac{\ln\left(\frac{C_0}{C_0-k}\right)}{\ln\left(\frac{C_0+k}{C_0}\right)} = 2 \tag{17}$$

Re-arranging terms we obtain:

$$\ln\left(\frac{C_0}{C_0-k}\right) = 2\,\ln\left(\frac{C_0+k}{C_0}\right). \tag{18}$$

Table 1 Main terms for studying deficiency aversion and loss aversion via the justice evaluation function

	Difference-based	Ratio-based
A. Deficiency Aversion		
General statement	$\lvert J^{\text{Deficiency}} \rvert > J^{\text{Excess}}$	
Mathematical statement	$\left\lvert \ln\left(\frac{C_0-k}{C_0}\right) \right\rvert > \ln\left(\frac{C_0+k}{C_0}\right)$	
Label	D^J	R^J
	$\left\lvert \ln\left(\frac{C_0-k}{C_0}\right) \right\rvert - \ln\left(\frac{C_0+k}{C_0}\right)$	$\frac{\left\lvert \ln\left(\frac{C_0-k}{C_0}\right) \right\rvert}{\ln\left(\frac{C_0+k}{C_0}\right)}$
	$\ln(C_0^2) - \ln(C_0^2 - k^2)$	$\frac{\ln\left(\frac{C_0}{C_0-k}\right)}{\ln\left(\frac{C_0+k}{C_0}\right)}$
Range	$0 < D^J < \infty$	$1 < R^J < \infty$
B. Loss Aversion		
General statement	$\lvert CJ^{\text{Loss}} \rvert > CJ^{\text{Gain}}$	
Mathematical statement	$\left\lvert \ln\left(\frac{A_0-k}{A_0}\right) \right\rvert > \ln\left(\frac{A_0+k}{A_0}\right)$	
Label	D^{CJ}	R^{CJ}
	$\left\lvert \ln\left(\frac{A_0-k}{A_0}\right) \right\rvert - \ln\left(\frac{A_0+k}{A_0}\right)$	$\frac{\left\lvert \ln\left(\frac{A_0-k}{A_0}\right) \right\rvert}{\ln\left(\frac{A_0+k}{A_0}\right)}$
	$\ln(A_0^2) - \ln(A_0^2 - k^2)$	$\frac{\ln\left(\frac{A_0}{A_0-k}\right)}{\ln\left(\frac{A_0+k}{A_0}\right)}$
Range	$0 < D^{CJ} < \infty$	$1 < R^{CJ} < \infty$

Notes Deficiency and excess are defined relative to the just reward C. Loss and gain are defined relative to the Time 1 actual reward A.

which leads to the following quadratic equation in k:

$$k^2 + C_0 k - C_0^2 = 0 \tag{19}$$

Solving Eq. (19), we find two real roots, one of them positive at:

$$k = C_0\left(\frac{\sqrt{5} - 1}{2}\right) \tag{20}$$

This root is quickly seen to include the Golden Number, approximately equal to 0.618. Thus, deficiency is felt twice as keenly as comparable excess when the deficiency (or excess) equals approximately 61.8% of the just reward.

It is extraordinarily pleasing to see the Golden Number appear in justice research, signaling again the deep involvement of nature in the sense of justice. The Golden Number joins logarithms and the beautiful number e in playing a part in the scientific description of the sense of justice, together with smaller but no less beautiful numbers and results. For example, the limit of the geometric mean in the distribution of relative ranks, which embeds two stalwarts—roots and factorials—approaches $1/e$:

$$\lim_{N\to\infty} \frac{\sqrt[N]{N!}}{N+1} = \frac{1}{e} \approx 0.368 \tag{21}$$

Thus, in a nonmaterialistic society that values one ordinal good and views justice as equality, as the population size increases, the average of the justice evaluation distribution moves leftward, attaining progressively larger absolute values of negative magnitudes and approaching its limit of −0.307 [64, p. 13][6]:

$$\lim_{N\to\infty} \ln\left(\frac{2\sqrt[N]{N!}}{N+1}\right) = \ln\left(\frac{2}{e}\right) = \ln(2) - 1 \approx -0.307 \tag{22}$$

Finally, the emergence of the Golden Number in the study of justice reinforces and illuminates the bond between justice and beauty [13–18]—a bond especially tight in languages like English, where the word "fair" signifies both "just" and "beautiful".[7]

Of course, the analysis shows that deficiency is felt twice as keenly as excess only for given special magnitudes of deficiency and excess (namely 0.618 of the just reward). Other magnitudes of deficiency and excess will yield other magnitudes of the ratio R^J. In general, the magnitude of R^J depends jointly on the magnitudes of the just reward C and the deficiency or excess k. Taking first partial derivatives of R^J with respect to C and k yields negative and positive quantities, respectively. Thus, the greater the just reward, the smaller the deficiency aversion coefficient, and the greater the amount of the deficiency or excess, the greater the deficiency aversion coefficient.

[6] This society belongs to a family of justice societies that satisfy two conditions, sometimes called the "Primitive alternatives" [52, 64 p. 9–10]: first, following Socrates in Plato's *Gorgias*, the members view justice as equality; and second, following Augustine's definition in *City of God*, they are " a people... bound together by a common agreement as to the objects of their love" and thus value the same goods or bads.

[7] Not all languages distinguish between "justice" and "fairness"—considered by Rawls [65] an important distinction. Thus, his phrase "justice as fairness" is rendered in French as "la justice comme équité" and in Spanish as "la justicia como imparcialidad." Both raise the new challenge of establishing the relation between the French "justice" and "équité" and between the Spanish "justicia" and "imparcialidad" and their relation to the Rawlsian English-language concepts of "justice" and "fairness" as well as "equity" and "impartiality".

5.2 Loss Aversion and the Golden Number

Paralleling the work above on deficiency aversion, we turn to the special case of loss aversion in which the ratio R^{CJ} equals 2. As discussed above, this case is important because of empirical evidence that loss is felt twice as keenly as comparable gain [62 p. 1288, 63 p. 1053–1054].

To analyze this case, we set the ratio R^{CJ} equal to 2 and solve for k. Paralleling exactly the work above on deficiency and excess, at the first step, we write:

$$\frac{\ln\left(\frac{A_0}{A_0-k}\right)}{\ln\left(\frac{A_0+k}{A_0}\right)} = 2 \tag{23}$$

Re-arranging terms we obtain:

$$\ln\left(\frac{A_0}{A_0-k}\right) = 2\ln\left(\frac{A_0+k}{A_0}\right) \tag{24}$$

which leads to the following quadratic equation in k:

$$k^2 + A_0 k - A_0^2 = 0 \tag{25}$$

Solving Eq. (25), we find two real roots, one of them positive at:

$$k = A_0\left(\frac{\sqrt{5}-1}{2}\right) \tag{26}$$

As before, this root is quickly seen to include the Golden Number, approximately equal to 0.618. Thus, loss is felt twice as keenly as comparable gain when the loss (or gain) equals approximately 61.8% of the actual reward at Time 1.

Also as before, the analysis shows that loss is felt twice as keenly as gain only for given special magnitudes of loss and gain (namely 0.618 of the actual reward). Other magnitudes of loss and gain will yield other magnitudes of the ratio R^{CJ}. In general, the magnitude of R^{CJ} depends jointly on the magnitudes of the actual reward A and the loss or gain k. Taking first partial derivatives of R^{CJ} with respect to A and k yields negative and positive quantities, respectively. Thus, the greater the Time 1 actual reward, the smaller the loss aversion coefficient, and the greater the amount of the loss or gain, the greater the loss aversion coefficient.

Table 2 reports a simple numerical example that illustrates these results. Panel A displays the justice evaluations and the loss aversion coefficient for three magnitudes of the Time 1 actual reward (80, 100, and 120), shown as three sets of columns, and four magnitudes of the loss or gain k (0, 25, 50, and 75), each displayed in a row. Thus, for all three levels of the Time 1 actual reward, when

k equals zero, there is neither loss nor gain, and the change in the justice evaluation CJ equals zero.

Each of the three sets of columns representing a magnitude of the Time 1 actual reward has five columns, two for the loss case, two for the gain case, and the loss aversion coefficient. For example, in the row for $k = 25$, looking at the set for the Time 1 actual reward equal to 80, the two columns for the loss case report the numeric representation of the log-ratio and the numerical approximation to CJ, and similarly for the gain case. Accordingly, the numeric log-ratio has a numerator of $(80 - 25 =)$ 55 in the loss case and $(80 + 25 =)$ 105 in the gain case, and a denominator of 80 in both cases. The numeric approximations to CJ are thus −0.375 in the loss case and 0.272 in the gain case. The loss aversion coefficient is then the absolute value of -0.375 divided by 0.272, or approximately 1.38.

As expected from the first partial derivative, the loss aversion coefficient declines as the Time 1 actual reward increases, diminishing in each row as the Time 1 actual reward increases from 80 to 100 to 120—for example, when $k = 25$, diminishing from 1.38 to 1.29 to 1.23. Similarly, the loss aversion coefficient increases as k increases, increasing in each column as k increases from 25 to 50 to 75—for example, when the Time 1 actual reward is 80, increasing from 1.38 to 2.02 to 4.19.

Table 2 also reports, in Panel B, the value of k when the loss aversion coefficient equals 2. As shown in Eq. (26), this value equals the product of the Golden Number and the Time 1 actual reward. Thus, the requisite values are approximately 49.4, 61.8, and 74.2 for the Time 1 actual rewards of 80, 100, and 120, respectively.

5.3 Remarks on Loss Aversion

As discussed above, the term "loss aversion" was introduced by Kahneman and Tversky [60, 61], who observed that "losses loom larger than gains" [60, p. 279]. Moreover, they found that losses are felt twice as strongly as gains [63, pp. 1053–1054]. Meanwhile, justice theory also predicts that losses are felt more strongly than gains, but the associated loss aversion coefficient is not constant and indeed takes on the value 2 only in the very special case involving the Golden Number.

In the spirit of this chapter, it is exciting to contrast the two approaches, which could be special cases of a larger framework. To that end, this section takes some steps to lay out explicitly the correspondence between them. First, both approaches are embedded in theory, specifically in prospect theory and justice theory. Second, the main driver in both approaches is a function, specifically the value function in prospect theory and the justice evaluation function in justice theory. Third, both the value function and the justice evaluation function emerged from empirical work. Fourth, both the value function and the justice evaluation function are reference-dependent. Fifth, in the value function gains or losses are assessed relative to a reference point, while in the justice evaluation function (i) deficiency and excess are assessed relative to the just reward and (ii) gain and loss are assessed relative to the actual reward at Time 1. Sixth, the outcomes (the value V and the justice evaluation J) range from negative numbers for the deficiency/loss condition,

Table 2 Loss aversion and loss aversion coefficient (LAC), by time 1 amount A_0 and loss/gain k

	$A_0 = 80$					$A_0 = 100$					$A_0 = 120$				
A. Change in the justice evaluation, by time 1 amount and loss or gain k, and loss aversion coefficient (LAC)															
k	$\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$	$\approx$	LAC	$\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$		LAC	$\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$	$\approx$	LAC
0	$\ln\left(\frac{80}{80}\right)$	0	$\ln\left(\frac{80}{80}\right)$	0	–	$\ln\left(\frac{100}{100}\right)$	0	$\ln\left(\frac{100}{100}\right)$	0	–	$\ln\left(\frac{120}{120}\right)$	0	$\ln\left(\frac{120}{120}\right)$	0	–
25	$\ln\left(\frac{55}{80}\right)$	-0.375	$\ln\left(\frac{105}{80}\right)$	0.272	1.38	$\ln\left(\frac{75}{100}\right)$	-0.288	$\ln\left(\frac{125}{100}\right)$	0.223	1.29	$\ln\left(\frac{95}{120}\right)$	-0.234	$\ln\left(\frac{145}{120}\right)$	0.189	1.23
50	$\ln\left(\frac{30}{80}\right)$	-0.981	$\ln\left(\frac{130}{80}\right)$	0.486	2.02	$\ln\left(\frac{50}{100}\right)$	-0.693	$\ln\left(\frac{150}{100}\right)$	0.405	1.71	$\ln\left(\frac{70}{120}\right)$	-0.539	$\ln\left(\frac{170}{120}\right)$	0.348	1.55
75	$\ln\left(\frac{5}{80}\right)$	-2.77	$\ln\left(\frac{155}{80}\right)$	0.661	4.19	$\ln\left(\frac{25}{100}\right)$	-1.39	$\ln\left(\frac{175}{100}\right)$	0.560	2.48	$\ln\left(\frac{45}{120}\right)$	-0.981	$\ln\left(\frac{195}{120}\right)$	0.486	2.02
B. Magnitude of k when loss aversion coefficient equals 2: $A_0\left(\frac{\sqrt{5}-1}{2}\right) \approx .618A_0$															
	≈ 49.4					≈ 61.8					≈ 74.2				

Note The loss aversion coefficient is defined as the ratio of the absolute value of the justice evaluation for a loss divided by the justice evaluation for a comparable gain

through zero for a neutral point in prospect theory and the point of perfect justice in justice theory, to positive numbers for the excess/gain condition.

To this point, there is a perfect correspondence between the two approaches. However, at first blush, the perfect correspondence seems to end, for the Kahneman and Tversky [61, p. S258-S259] losses and gains are represented by negative and positive numbers, respectively, while the actual rewards embedding both deficiency and excess are represented by positive numbers. Thus, the graph of the value function occupies Quadrants I and III in the Cartesian plane, while the graph of the justice evaluation function occupies Quadrants I and IV. Moreover, a hallmark of the value function is that it is S-shaped, with the response to losses convex and the response to gains concave, while the justice evaluation function is concave throughout.

However, it may be possible for the JEF to approximate the value function by a simple procedure: treat losses as bads and gains as goods, and represent bads by negative numbers.[8] The resulting graph is an S-shaped curve. But the approximation is not complete, for the S-shaped graph of the value function is asymmetric, with greater steepness for losses than for gains, while the S-shaped graph of the transformed justice evaluation function is symmetric. Put differently, the asymmetry of the value function ensures that losses loom larger than gains; but the symmetric transformed JEF loses loss aversion.

Can loss aversion be restored to the transformed JEF? Recall from Sect. 2 that justice theory distinguishes between the experienced justice evaluation and the expressed justice evaluation, via the expressiveness coefficient, the absolute value of the Signature Constant θ in Eq. (1). If losses elicit greater expressiveness than gains (more shouting, say, or less whispering), then the transformed justice evaluation function becomes asymmetric and can approximate the value function. This avenue has a further advantage, namely, incorporating the expressiveness

[8] Framing is powerful, and it is possible that a loss is framed as receiving a bad rather than a smaller amount of a good.

coefficient renders the loss aversion coefficient constant. It can be constant at two simply by setting the expressiveness coefficient to two. However, the asymmetric transformed JEF loses the connection to the Golden Number.

To see this way of attempting to bring the JEF into alignment with the value function, Table 3 provides a numerical example and Fig. 3 provides visualization. The basic numerical example is taken from the middle set of justice evaluations in Table 2, namely, the set in which the Time 1 amount is 100. This set of five columns becomes the leftmost set in Table 3. The second and third sets of five columns each present the symmetric and asymmetric transformed justice evaluations, respectively. In both sets, losses are treated as bads and gains as goods, and in both sets, losses are represented by negative numbers. The two sets differ, however, in the expressiveness coefficient for losses, set at one in the second set and two in the third (it could, of course, be any other positive number, but two has the advantage that it corresponds to the Tversky and Kahneman empirical finding).

As shown in Table 3, the original justice evaluations in the leftmost set display the loss aversion inherent in the justice evaluation function. However, there is no loss aversion in the middle set of figures corresponding to the symmetric transformed JEF. Moreover, in the rightmost set of figures corresponding to the asymmetric transformed JEF, loss aversion is introduced via the expressiveness coefficient and is thus a constant, losing the connection to the Golden Number.

Figure 3 provides a visualization of the original and transformed justice evaluations. As shown, Panels A and B depict the figures in the first set of columns in Table 3. They differ, however, in that in Panel A the JEF is shown in the original coordinates, while in Panel B it has been translated leftward so that losses are represented by negative numbers and the loss or gain is no longer relative to the Time 1 amount of 100 but rather to the zero point between losses and gains.

Panels C and D correspond to the second and third sets of figures in Table 3. In both, losses are treated as bads and goods as gains. There is no loss aversion in Panel C, which depicts the symmetric transformed JEF. Meanwhile, the loss aversion in Panel D, which depicts the asymmetric transformed JEF, is constant, with the loss aversion coefficient fixed at two, and importantly, the link to the Golden Number is lost.

Table 3 Loss aversion and loss aversion coefficient (LAC), by time 1 amount A_0 and loss or gain k: Unifying the prospect theory and justice theory approaches

	Losses and gains treated as goods $k > 0$ for both losses and gains $A_0 = 100$					Losses treated as bads, gains as goods $\theta = -1$ for bads, $+1$ for goods $k < 0$ for losses, $k > 0$ for gains $A_0 = 100$					Losses treated as bads, gains as goods $\theta = -2$ for bads, $+1$ for goods $k < 0$ for losses, $k > 0$ for gains $A_0 = 100$				
k	$\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$	$\approx$	LAC	$-\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$		LAC	$-2\ln\left(\frac{A_0-k}{A_0}\right)$	$\approx$	$\ln\left(\frac{A_0+k}{A_0}\right)$	$\approx$	LAC
0	$\ln\left(\frac{100}{100}\right)$	0	$\ln\left(\frac{100}{100}\right)$	0	–	$-\ln\left(\frac{100}{100}\right)$	0	$\ln\left(\frac{100}{100}\right)$	0	–	$-2\ln\left(\frac{100}{100}\right)$	0	$\ln\left(\frac{100}{100}\right)$	0	–
25	$\ln\left(\frac{75}{100}\right)$	−0.288	$\ln\left(\frac{125}{100}\right)$	0.223	1.29	$-\ln\left(\frac{125}{100}\right)$	−0.223	$\ln\left(\frac{125}{100}\right)$	0.223	1	$-2\ln\left(\frac{125}{100}\right)$	−0.446	$\ln\left(\frac{125}{100}\right)$	0.223	2
50	$\ln\left(\frac{50}{100}\right)$	−0.693	$\ln\left(\frac{150}{100}\right)$	0.405	1.71	$-\ln\left(\frac{150}{100}\right)$	−0.405	$\ln\left(\frac{150}{100}\right)$	0.405	1	$-2\ln\left(\frac{150}{100}\right)$	−0.810	$\ln\left(\frac{150}{100}\right)$	0.405	2
75	$\ln\left(\frac{25}{100}\right)$	−1.39	$\ln\left(\frac{175}{100}\right)$	0.560	2.48	$-\ln\left(\frac{175}{100}\right)$	−0.560	$\ln\left(\frac{175}{100}\right)$	0.560	1	$-2\ln\left(\frac{175}{100}\right)$	−1.12	$\ln\left(\frac{175}{100}\right)$	0.560	2

Note The loss aversion coefficient is defined as the ratio of the absolute value of the justice evaluation for a loss divided by the justice evaluation for a comparable gain

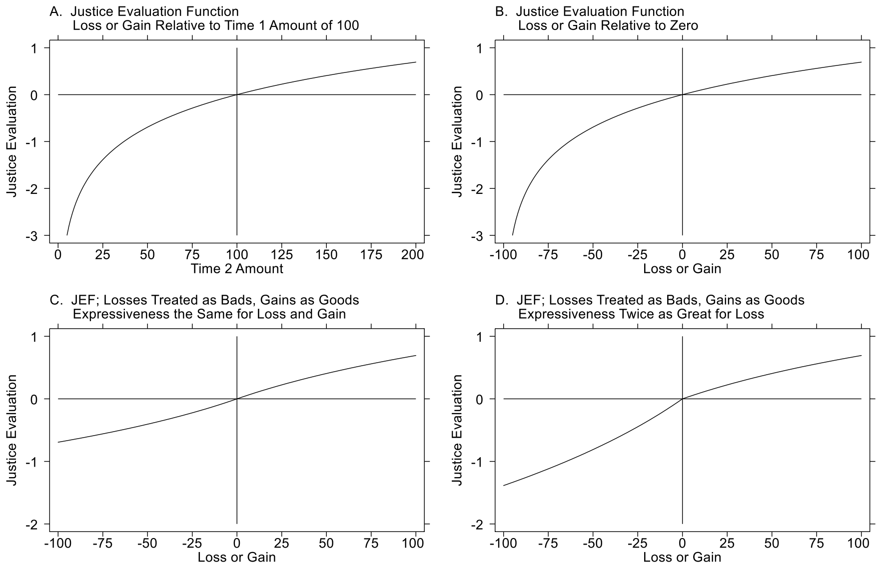


Fig. 3 Justice evaluation function approximating the value function

As can be readily appreciated, the stage is set for further research in several directions, inclusive of understanding more deeply the underlying differences between the loss aversion process in the value function and in the justice evaluation function as well as in the two (symmetric and asymmetric) transformed justice evaluation functions. As well, further research can assess whether and how researchers and respondents may be unconsciously replicating the "divine proportions" of the Golden Number.

6 A Few Words on Money, Games, and Couples

Recently, several scholars have suggested that the special beauty of the Golden Number proportions may extend to matters of allocation and distribution, including salary schedules and classic games like the Ultimatum Game and the Common Pool Resource Game [15–17, 66–70]. Langen [67, 68] observes that salary schedules in hierarchical organizations might replicate Golden Number proportions in the progression from step to step. Thus, salaries in civil service, military service, private firms, public corporations, athletic teams, and dramatic and musical performing ensembles, may usefully be examined for evidence or traces of the Golden Number. Similarly, the many historical accounts of distribution—such as the national systems for allocating prize money from the capture of enemy ships in naval warfare—provide rich territory for assessing actual and preferred proportions.

In the realm of dyads, Langen [66–68] suggested that the Golden Number might be a solution to the Ultimatum Game, and Schuster [15] and Suleiman [16], working independently and using different approaches, both obtained the Golden Number as a new solution for the Ultimatum Game, with a division of 0.618 for the proposer and 0.382 for the responder. Schuster [15] used ideas of optimality and infinite continued fractions to derive the solution at the Golden Number. Suleiman [16] proposed a model of "economic harmony" in which utility is defined as a function of the ratio between actual and aspired payoffs and then derived the solution at the Golden Number. Suleiman [17] subsequently extended the argument to bargaining games with alternating offers, again finding the Golden Number as the solution.

Meanwhile, Vermunt [70] suggested that a fair allocation to self and other may lie between equal division and Golden Number division, with 0.618 for self and 0.382 for other. Of course, for a votary of both Francis of Assisi and the Golden Number, the division might be 0.382 for self and 0.618 for other. This possibility that fair allocation may lie between equal division and Golden Number division may be interpreted—or occur—in several ways. First, the observer's idea of the just reward may be *a specific number* between 0.382 and 0.5 or between 0.5 and 0.618, depending on whether the observer's selection for the rewardee (self or other) is the smaller or the larger. Second, the observer's idea of the just reward may be *any number* between 0.382 and 0.5 or between 0.5 and 0.618, generating a justice zone. More generally, a third-party observer's ideas of the just reward for the two prospective recipients could be either (1) a specific set of complementary proportions between 0.382–0.618 and 0.618–0.382 (e.g., 0.4 & 0.6, 0.5 & 0.5, 0.6 & 0.4, etc.), or (2) any set of complementary proportions between 0.382–0.618 and 0.618–0.382.

Now consider couples of a special kind, romantic couples and married couples. How might they react to various allocations, including both allocations made inside and outside the dyad? Imagine a couple where the partners are close to identical in age, ability, education, aspirations, and so on. The two partners think they merit equal salaries so that each partner's justice evaluation equals the log of the ratio of that partner's earnings to the average of both partners' earnings. However, for unexplained reasons, the bride earns more than the groom. Thus, the bride has a positive justice evaluation and the groom a negative one.

Developing the link between individuals' justice evaluations and cohesiveness (in both dyads and larger groups), justice theory embeds the idea, based on Aristotle's notion that love is possible only to equals, that cohesiveness requires at the very least what may be called an equality fantasy. Accordingly, the partners' cohesiveness declines with disparity in their justice evaluations. Formally, the couple's cohesiveness reduces to the logarithm of the ratio of the smaller to the larger earnings. Thus, cohesiveness is lowest when the ratio of the smaller to the larger earnings is very small and reaches its maximum when the two earnings amounts are equal.

Justice theory also derives the cohesiveness for dyads in which neither partner has earnings, or only one partner has earnings, based on the idea that in such case, the valued good used to drive their self-worth is ordinal. The ordinal-case cohesiveness is the logarithm of one-half.

A number of testable predictions follow immediately, concerning, for example, whether a marriage is strengthened, weakened, or left unaltered if one partner takes up employment or becomes unemployed or retires, etc. The general prediction is that if the spouses' earnings ratio exceeds one-half, marital cohesiveness is greater when both partners are employed, but if their earnings ratio is less than one-half, marital cohesiveness is lower when both partners are employed.

Returning to the special case in which the two partners' salaries are unequal, the couple's cohesiveness equals the logarithm of the ratio of the smaller to the larger earnings. If the earnings ratio assumes the Golden Number proportions, the ratio of the smaller to the larger earnings is approximately 0.618. Thus, it lies between one-half (the ratio when only one partner is employed) and one (when the two salaries are equal). Accordingly, the Golden Number salary ratio is within a "stable" zone, as it does not lead the couple to alter their employment situations.

7 Conclusion

- The core story of this chapter asked a simple question—When is deficiency (loss) felt twice as keenly as comparable excess (gain)? The answer—When the deficiency (loss) or excess (gain) are approximately 0.618 of the just amount (Time 1 amount)—was a great surprise. For 0.618 is the Golden Number, the "divine proportion" in mathematics. Thus, the core story exemplifies the beauty, unity, and surprise of mathematics, science, and art.
- The chapter went on to consider whether and how the key driver in this appearance of the Golden Number, namely, the justice evaluation function of justice theory, can approximate the value function of prospect theory.
- The chapter also considered additional exciting recent developments about the Golden Number in social science, specifically in games and dyads.
- Embedded in the work reported in this chapter are several directions for future research. For example, the fact that empirical results indicate that loss is felt twice as keenly as gain suggests that researchers and/or respondents may be (unconsciously) choosing magnitudes of loss and gain that hover about the Golden Number, that is, themselves in thrall to this magic quantity.
- This work raises the question, what is the connection between loss and gain, on the one hand, and bads and goods, on the other?
- A further new question is, How is this connection itself linked to the choice of numbers that both ordinary people and researchers use "in their minds" to represent loss and gain, bads and goods?

- To this mix, we add one final set of questions: What does it mean when the Golden Number suddenly appears in scientific work? Is it a signal from nature? To be sure, it brings beauty, and it brings surprise, and it certainly serves to link mathematics, science, and art. But is there something else?

Core Messages

- Mathematics, science, and art share three hallmarks: beauty, unity, surprise.
- The Golden Number of mathematics exemplifies the three hallmarks, both internally and in its relations with science and art.
- The justice evaluation function of social science exemplifies the three hallmarks, both internally and in its relations with mathematics.
- The Golden Number answers the answer to a simple question arising from the justice evaluation function.
- Is the Golden Number a signal from nature?

Acknowledgements The author presented earlier versions of portions of this paper at the National Science Foundation, Washington, DC, March 2003; at Zhejiang University, Hangzhou, China, April 2004; at Tilburg University, the Netherlands, June 2004; at the annual meeting of the Swedish Sociological Society, Skovde, Sweden, February 2005; at the EITM Summer Institute, University of Michigan, June 2006; at the Annual Conference of EQUALSOC, the European Commission's Network of Excellence on Economic Change, Quality of Life, and Social Cohesion, Barcelona, Spain, September 2006; at the University of Kiel, May 2011; at the University of Stockholm, Sweden, August 2012; at the 14th Biennial Conference of the International Society for Justice Research, College of Management, Rishon LeZion, Israel, September 2012; at the Institute for the Study of Labor (IZA), Bonn, Germany, March 2016; and at the EITM Summer Institute, Hobby School of Public Affairs, University of Houston, Houston, Texas, June 2017. I am grateful to participants at those meetings and especially to Richard Alba, Norman Bradburn, Deborah Frisch, James S. Granato, Samuel Kotz, Eva Meyersson Milgrom, Dahlia Moore, Clara Sabbagh, Stefan Schuster, Frank P. Scioli, Ramzi Suleiman, Elizabeth Thomson, Kjell Y. Törnblom, and Murray Webster for many valuable comments and suggestions. I also gratefully acknowledge the intellectual and financial support provided by New York University.

References

1. Einstein A ([1879–1902]/2020) The collected papers of Albert Einstein. Volume 1: the early years, 1879–1902 (English translation supplement). Princeton University Press, Princeton, pp 165–166. Digital project in process. https://einsteinpapers.press.princeton.edu/
2. Wilson EO (1998) Consilience: the unity of knowledge. Knopf, New York
3. Thompson SP ([1910]/1946) Calculus made easy, 3rd edn. St. Martin's Press, New York
4. Fararo TJ (1989) The spirit of unification in sociological theory. Sociol Theory 7(2):175–190
5. Fararo TJ, Skvoretz J (1987) Unification research programs: integrating two structural theories. Am J Sociol 92(5):1183–1209. https://doi.org/10.1086/228632

6. Fararo TJ, Skvoretz J (2002) In: Berger J, Zelditch M (eds) New directions in contemporary sociological theory. Rowman and Littlefield, Boulder, Colorado, pp 295–316
7. Franzosi R (2004) From words to numbers: narrative, data, and social science. Cambridge, New York
8. Jasso G (2007) Theoretical unification in justice and beyond. Soc Justice Res 20(3):336–371. https://doi.org/10.1007/s11211-007-0055-7
9. Törnblom KY, Jasso G, Vermunt R (2007) Theoretical integration and unification: a focus on justice. Soc Justice Res 20(3):263–269. https://doi.org/10.1007/s11211-007-0044-x
10. Wagner D, Berger J (1985) Do sociological theories grow? Am J Sociol 90(4):697–728. https://doi.org/10.1086/228142
11. Cascardi AJ (2000) Two kinds of knowing in Plato, Cervantes, and Aristotle. Philos Lit 24 (2):406–423. https://doi.org/10.1353/phl.2000.0030
12. Ortega y Gasset J (1961) Meditations on Quixote. Translated by Evelyn Rugg and Diego Marín. W. W. Norton, New York. (Original work published 1914)
13. Livio M (2002) The golden ratio: the story of phi, the world's most astonishing number. Broadway Books, New York
14. Scarry E (1999) On beauty and being just. Princeton University Press, Princeton
15. Schuster S (2017) A new solution concept for ultimatum game leading to the golden ratio. Sci Rep. https://doi.org/10.1038/s41598-017-05122-5
16. Suleiman R (2017) Economic harmony: an epistemic theory of economic interactions. Games 8:2. https://doi.org/10.3390/g8010002
17. Suleiman R (2018) On gamesmen and fair men: explaining fairness in non-cooperative bargaining games. Roy Soc Open Sci 5:171709. https://doi.org/10.1098/rsos.171709
18. Wang T, Mo L, Mo C, Li HT, Cant JS, Zhong L, Cupchik G (2015) Is moral beauty different from facial beauty? Evidence from an fMRI study. Soc Cogn Affect Neurosci 10(6):814–823. https://doi.org/10.1093/scan/nsu123
19. Babbitt N (2018) Barking with the big dogs: on writing and reading books for children. Palgrave Macmillan, New York
20. Scalia A (1989) The rule of law as a law of rules. Univ Chicago Law Rev 56(4):1175–1188
21. Adams JS, Freedman S (1976) Equity theory revisited: comments and annotated bibliography. In: Berkowitz L, Walster E (eds) Advances in experimental social psychology, vol 9. Academic, New York, pp 43–90
22. Berkowitz L, Walster E (1976) Preface. In: Berkowitz L, Walster E (eds) Advances in experimental social psychology, vol 9. Academic, New York, pp xi–xii
23. Homans GC (1976) Commentary. In: Berkowitz L, Walster E (eds) Advances in experimental social psychology, vol 9. Academic, New York, pp 231–244
24. Walster E, Berscheid E, Walster GW (1973) New directions in equity research. J Pers Soc Psychol 25(2):151–176. https://doi.org/10.1037/h0033967
25. Jasso G (2008) A new unified theory of sociobehavioral forces. Eur Sociol Rev 24(4):411–434. https://doi.org/10.1093/esr/jcn023
26. Jasso G (2017) Inequality in the distribution of a good is a bad, and inequality in the distribution of a bad is a good. Eur Sociol Rev 33(4):604–614. https://doi.org/10.1093/esr/jcx062
27. Rayo L, Becker GS (2007) Evolutionary efficiency and happiness. J Polit Econ 115(2):302–337. https://doi.org/10.1086/516737
28. Jasso G (2015) Thinking, saying, doing in the world of distributive justice. Soc Justice Res 28 (4):435–478. https://doi.org/10.1007/s11211-015-0257-3
29. Jasso G, Törnblom, KY, Sabbagh C (2016) Distributive justice. In: Sabbagh C, Schmitt M (eds) Handbook of social justice theory and research. Springer, New York, pp 201–218. https://doi.org/10.1007/978-1-4939-3216-0_11
30. Jasso G, Wegener B (1997) Methods for empirical justice analysis: part I. Framework, models, and quantities. Soc Justice Res 10(4):393–430. https://doi.org/10.1007/BF02683292

31. Jasso G (1978) On the justice of earnings: a new specification of the justice evaluation function. Am J Sociol 83(6):1398–1419. https://doi.org/10.1086/226706
32. Wegener B, Steinmann S (1995) Justice psychophysics in the real world: comparing income justice and income satisfaction in East and West Germany. In: Kluegel J, Mason D, Wegener B (eds) Social justice and political change: public opinion in capitalist and post-communist states. Aldine de Gruyter, New York, pp 151–175
33. Brickman P, Folger R, Goode E, Schul Y (1981) Micro and macro justice. In: Lerner MJ, Lerner SC (eds) The justice motive in social behavior. Plenum, New York, pp 173–202
34. Friedman M (1977) Fair versus free. Newsweek (4 July 1977, p. 70). Originally delivered as a commencement address at Virginia Polytechnic Institute, 28 May 1977
35. Jasso G (1999) How much injustice is there in the world? Two new justice indexes. Am Sociol Rev 64(1):133–168. https://doi.org/10.2307/2657282
36. Jasso G (1990) Methods for the theoretical and empirical analysis of comparison processes. Sociol Methodol 20:369–419. https://doi.org/10.2307/271091
37. Jasso G (1993) Choice and emotion in comparison theory. Ration Soc 5(2):231–274. https://doi.org/10.1177/1043463193005002006
38. Sheehan N (1991) After the war was over: Hanoi and Saigon. Random House, New York
39. Bellou A (2017) Male wage inequality and marital dissolution: Is there a link? Can J Econ 50 (1):40–71. https://doi.org/10.1111/caje.12250
40. Bernoulli D ([1738]/1954) Exposition of a new theory on the measurement of risk. Econometrica 22:23–36. Translated from the Latin by L. Sommer, Specimen theoriae novae de mensura sortis
41. Fechner GT ([1860]/1907) Elemente der psychophysik, vol 2. Leipzig, Breitkopf and Härtel
42. Fechner G ([1877]/1968) In sachen der psychophysik. Leipzig, Breitkopf and Härtel. (edn published by Bonset EJ, Amsterdam)
43. Jasso G (1996) Exploring the reciprocal relations between theoretical and empirical work: the case of the justice evaluation function (paper in honor of Robert K. Merton). Sociol Methods Res 24(3):253–303. https://doi.org/10.1177/0049124196024003002
44. Jasso G (2006) Homans and the study of justice. In: Treviño AJ (ed), George C. Homans: History, theory, and method. Paradigm, Boulder, pp 203–227
45. van den Bos K, Cropanzano R, Kirk J, Jasso G, Okimoto T (2015) Expanding the horizons of social justice research: three essays on justice theory. Soc Justice Res 28(2):229–246. https://doi.org/10.1007/s11211-015-0237-7
46. Jasso G (2008) The terms and relations of comparison, referential, and relative processes. In: Diekmann A, Eichner E, Schmidt P, Voss T (eds) Rational choice: theoretische analyse und empirische resultate. Festschrift für Karl-Dieter Opp. Wiesbaden, Germany: VS Verlag, Wiesbaden pp 265–290. https://doi.org/10.1007/978-3-531-90866-3_15
47. Marx K ([1891]/1849) 1968 Wage labour and capital. In: Marx K, Engels F selected works
48. James W [(1891)/1952] The principles of psychology Britannica Chicago
49. Falk A, Knell M (2004) Choosing the Joneses: endogenous goals and reference standards. Scand J Econ 106(3):417–435. https://doi.org/10.1111/j.0347-0520.2004.00370.x
50. Jasso G (1989) The theory of the distributive-justice force in human affairs: analyzing the three central questions. In: Berger J, Zelditch M, Anderson B (eds) Sociological theories in progress: new formulations. Sage, Newbury Park, California, pp 354–387
51. Layard R (2003) Happiness: has social science a clue? Lionel Robbins memorial lectures. Centre for economic performance, London
52. Plato ([4th century BCE]/1952) The dialogues of Plato. Translated by Benjamin Jowett. Britannica, Chicago
53. Huntley HE (1970) The divine proportion: a study in mathematical beauty. Dover, New York
54. Struik DJ (1987) A concise history of mathematics, 4th edn. Dover, New York
55. Euclid ([circa 300 BCE]/1952) The thirteen books of Euclid's elements. Translated by Sir Thomas L. Heath. Britannica, Chicago

56. Bronshtein IN, Semendyayev KA (1985) Handbook of mathematics. English translation edited by Hirsch KA. Van Nostrand Reinhold, New York. [ISBN: 9780442211714]
57. Benavoli A, Chisci L, Farina A (2009) Fibonacci sequence, golden section, Kalman filter and optimal control. Signal Process 89(8):1483–1488. https://doi.org/10.1016/j.sigpro.2009.02.003
58. Nematollahi AF, Rahiminejad A, Vahidi B (2020) A novel meta-heuristic optimization method based on golden ratio in nature. Soft Comput 24:1117–1151. https://doi.org/10.1007/s00500-019-03949-w
59. Adams JS (1963) Toward an understanding of inequity. J Abnorm Soc Psychol 67(5):422–436. https://doi.org/10.1037/h0040968
60. Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–292. https://doi.org/10.2307/1914185
61. Tversky A, Kahneman D (1986) Rational choice and the framing of decisions. J Bus 59(4): S251–S278
62. Rabin M (2000) Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68(5):1281–1292. https://doi.org/10.1111/1468-0262.00158
63. Tversky A, Kahneman D (1991) Loss aversion in riskless choice: a reference-dependent model. Quart J Econ 106(4):1039–1061. https://doi.org/10.2307/2937956
64. Jasso G (1980) A new theory of distributive justice. Am Sociol Rev 45(1):3–32. https://doi.org/10.2307/2095239
65. Rawls J (1958) Justice as fairness. Philos Rev 67(2):164–194
66. Langen M (2006) Konkurrenzlos gut (Good without competition). ChangeX 2006, article 2502. http://www.changex.de/Article/article_2502
67. Langen M (2006) Zum design eines fairen Gehaltssystems (On the design of a fair salary system). https://www.knowledge-communities.com/downloads/FairesGehaltssystem.pdf
68. Langen M (2007) Der goldene Schnitt (The golden section). Brand eins 7:146–147. https://www.brandeins.de/magazine/brand-eins-wirtschaftsmagazin/2007/zu-viel/der-goldene-schnitt
69. Suleiman R (2014) What is a fair salary? http://vixra.org/pdf/1409.0075v6.pdf
70. Vermunt R (2014) The good, the bad, and the just: How modern men shape their world. Ashgate, Surrey, England

Guillermina Jasso (Ph.D., Johns Hopkins) studies fairness, inequality, and international migration, together with methods for their theoretical and empirical analysis. Her contributions include a formula for fairness assessment, a formula for decomposing overall injustice into injustice due to poverty and injustice due to inequality, linking inequality between persons and inequality between subgroups, two families of probability distributions, and a procedure for devising a point system for immigrant selection. Jasso has served as Chair of the Social Statistics Section of the American Statistical Association and six Sections of the American Sociological Association and as member and Chair of the US Census Scientific Advisory Committee. She is an elected member of the Johns Hopkins Society of Scholars, a Fellow of the American Association for the Advancement of Science, and was a Fellow at the Center for Advanced Study in the Behavioral Sciences. In 2015 Jasso won the Paul F. Lazarsfeld Award, given by the American Sociological Association's Methodology Section for distinguished contributions to sociological methodology. Jasso's Erdős number is 3.

Integration of Film Theory and Studies with Neuroscience

5

Miloš Milošević and Branko Sujić

The interpretation of Dreams is the royal road to a knowledge of the unconscious activities of the mind.

Sigmund Freud [1]

Summary

The purpose of this paper is to recommend and analyze the possibility and potential of integrating research in neuroscience and film (and media) theory. The starting point is the widespread use of oneiric metaphors in film studies, which make up the base of these primarily psychoanalytic theories and methods in its research. The history and postulates of the theoretical application of the comparison of film and dreams in film theory develop as are some of the new neuroscience findings that examine the neurophysiologic basis of dreaming and film reception. Based on the theories and empirical data presented, it has been demonstrated that using the method of comparing films and dreams in film study is justified and empirically grounded as a neurophysiologic similarity between the two processes exists. Heightened activity of the visual cortex and the inhibition of parts of the prefrontal cortex, which lead to loss of self-representation, indicate the two processes' main similarity. The main difference

M. Milošević (✉)
Singidunum University, Danijelova 32, 11000 Belgrade, Serbia

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Belgrade, Serbia

B. Sujić
Department of Dramatic Arts, University of Arts, Belgrade, Serbia

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_5

is the subject's presence in an imaginary world and a degree of authorship. There is a need for further study to comprehend the process of secondary elaboration. Theoretical bases for further research came along with the technological development of virtual reality media in the direction of the post-screen image.

Scan the QR code above to have "A Look Beyond the Foreseeable—the boundary between still and moving images". (Adapted from https://www.behance.net/gallery/78717387/A-Look-Beyond-the-Foreseeable).

The code of this chapter is 01010110 01010010.

1 Introduction

Regardless of the perception of the nature of human knowledge or its relationship toward the truth, creating an all-encompassing system of knowledge can be singled out as the most fundamental goal of science and scientific research. It can be also be considered an ideal of science and scientific research. Suppose these postulated goals and the ideals of science, that is, its purpose and mission in human society, are considered. In that case, the division of science into fields and disciplines primarily reflects the limits of human knowledge capacity and ability than the nature of scientific knowledge and the subject of its study. From this perspective, an integrative and interdisciplinary approach toward scientific research appears to be the logical-even necessary-step toward overcoming the limits mentioned above of human knowledge capacities and abilities for the most comprehensive achievement of the postulated goals.

Even though it is difficult today to find members of the scientific community who are against interdisciplinary approach toward scientific research, or those who have never vouched for it in an application for funds and grants, what frequently happens in practice is the phenomenon that "interdisciplinary researches" imply the propagation of one discipline over the research field of other disciplines without the sincere appreciation of the singularities of their methods and interests. Multiple "specialized" editions of scientific journals exist in which the "interdisciplinary" approach studies an occurrence which is not specific as a subject of study of the said journal without any methodological or theoretical qualitative step forward which the integration of several disciplines would necessarily be the result of. Similarly, the interdisciplinary approach is a good excuse for scientific conferences for those fields where there is not enough interest to call upon them and research from other fields, while the dialog of various disciplines and approaches in research, as well as changes in research practice that could be the consequences of said dialogs, as a result, are of secondary importance. Contrariwise, accelerated technological progress and radical social changes must require an interdisciplinary approach from science so that it might be able to catch up with the speeding changes meant to describe. Thusly, for example, the consequences of the accelerated development of new technologies may reflect the fact that theory is falling behind in practice. Often, a new technological observation is already obsolete at the moment of its publication. It happens because the subject of study is obsolete technologies and platforms that were current at the time of beginning the research [2]. The theory of and study of the film (and media), because of the combination of the exceptionally technical nature of its own subject of study and exceptionally theoretical and speculative methodology, is one of the scientific disciplines in which this problem is exceptionally complex. Therefore, an interdisciplinary approach appears to be the only possible solution [3].

In turn, film theory would influence numerous other social and humanistic theories, schools of thought, and scientific disciplines, such as psychology, sociology, psychoanalysis, and similar. Therefore, film theory is a discipline understood as extremely open when it comes to the interdisciplinary approach to research. On the other hand, a profound analysis of the content of periodicals, technical literature, and curricula of studies is not necessary to observe an almost complete absence of applied quantitative and experimental methods of research on film studies, theater, and dramatic arts [4, 5]. The reason for this can be due to the widespread thinking of researchers that the subject of study of these sciences is so specific that it cannot be quantified appropriately, nor researched using quantitative methods [6], as well as that each attempt at quantification leads, in fact, to trivializing the problem. Still, if the growing field of quantitative studies of culture through Social Computing and Digital Humanities comes into play [7], applying quantitative methods of research integrated with the traditional natural, medical, and technical–technological fields appears as a possibility that we cannot just reject it outright. The title of this chapter has embedded one of the directions of this possible integration.

Integrating film study and neuroscience not only would express the need for new technologies formed by social circumstances but acts as a natural continuation of current film study to this date [8, 9], even though numerous researchers from this field have outright rejected such assertions. Namely, from the film's very origin and its theoretical deliberation, films and dreams appeared to share a similarity. The consequence of this has been the great use of psychoanalytical assumptions in the film study [10–13]. Today's link between films and dreams is too narrow to cover the entire range of intertwining film theory and psychoanalysis [12]. However, it represents the first strong point that emerged from the two fields merged as one, justifying the psychoanalytic method and terminologies in film theory [13, 14]. Psychoanalytic theories help the foundation of approaches toward studying film, from feminism to viewers' hermeneutics [12]. Regardless of the described influence and numerous results of the application of psychoanalysis in studying film [11, 12], the question that arises is whether it is justified to use metaphors and analogies of the film-dream based on visible and assumed similarities, that is, whether it is empirically valid. Technologically advanced, methodologically expanded, compatible, and open to an interdisciplinary approach, neuroscience research could answer this question.

To avoid the earlier described one-sidedness of the practice of interdisciplinary studies, a question raised should be at the very start of why should neurosciences be interested in film theory at all; that is, if film studies can already advance using the methods and insights of neurosciences for validating their own theoretical assumptions, then what can film study offer to neuroscience? Suppose the primary goal of neuroscience is to create a connection between behavior and the central nervous system [15], to explain the neurological foundation and neurological mechanisms behind the behavior. In that case, the answer appears on its own. To realize its goal, that is, to successfully interpret the data obtained from the central nervous system, neuroscience needs a theoretical model of behavior, which can be found, among other places, in film study. These theoretical models are qualitatively different from those offered by other social sciences and humanities, such as psychology or sociology, and undoubtedly essential if the goal is to achieve self-proclaimed all-comprehension of knowledge from the start. That is how this connection is neither one-sided nor is integrating the use of neuroscience recommended only to develop film study. Still, rather film study can equally assist in the development of neuroscience.

Further elaboration, evidence, and illustrations of its justification of the postulated hypotheses linking the justification for the deeper connection and integration of neuroscience and film study present the primary goal. To this end, an analysis of the justification for using metaphors and analogies for film-dream and its results brought forth of its use in film theory will first take place. It will be done so by comparing the results thus far of research on film theory on the observed and assumed similarities of the two processes and the results of neurological findings, that is, the results of research on the neurophysiological basis of dreaming and neurophysiological basis of film reception. Namely, the first step consists of a brief history and postulation of the film compared with the dream in film theory. It includes some of the new neuroscience findings that deal with the neurophysiological basis of sleeping and film reception. The second step will critically consider the theoretical postulates of film theory from both the aspects of empirical grounding and all-encompassing applied theoretical solutions. Finally, the authors draw perspectives and implications to indicate the potential for integrative research toward neuroscience and film theory.

2 Film and Dream in Film Theory

Numerous authors have declared the similarity between films and dreams [10–14]. The use of metaphor and analogy originating out of noticed similarities is nearly as old as the art of film [10, 13]. The oneiric characteristic of the film was already recognizable by the pioneers of film theory, such as Hugo Münsterberg, Ricciotto Canudo i Jean Epstein [10–13]. Throughout the 1930s, American film and film theory have become applicable to the psychoanalytic method. The film analyzed by Sigmund Freud's method of analyzing dreams [1] insists on recognizing the mechanisms of pleasure, subconscious impulses with the final goal of understanding social relationships, and their pathologies. The key point for justifying this approach is the metaphor of the film-dream [14].

Following World War II, this theme is updated in film theory within the framework of the systematic study of film situation [11]. Serge Lebovici demonstrates the similarity between film and dream as an expressive means based on visual nature, absence of concrete temporal and spatial connections, and strict causal principles, mutual to film and the language of dream [16]. Hugo Mauerhofer finds similarities between films and dream situations in the passivity, anonymity, and reception of the viewer (dreamer), which leads to a retreat from reality while mentioning that the key difference is that the dreamer produces his own dreams, while the film viewer receives it as a finished product [17]. Like Mauerhofer, Lebovici demonstrates the similarity of the film viewer's position and dreamer in the second step of demonstrating the similarities between film and dream. A conclusion is that the reception of film and dream are similar psychological processes [16]. Susanne Langer mentions that how a film is presented is like a dream because it creates a virtual present, a framework of directorial representation leading to the conclusion that the film viewer takes the dreamer's position [18].

Edgar Morin [19] stresses that in both film reception and dreaming and objective perception, robust subjective mechanisms participate, which induce a state of infantile regression and overwhelm the subject. As a similarity, he adds that film picture, just as in dreaming, does not represent the true presence of objects, but rather their doubles, given that film picture is still material.

Observing the materiality of the film pictures and screen, Jean Mitry notes that, due to its distance from the object, the film picture possesses a mental picture [20], not merely perceptive ones. Mitry defines the film experience as a para-oneiric state. He places it somewhere between the daydream and dreams, based on the intensity and degree of replacing reality with the imaginary, and conscious that it deals with the imaginary and not real experience [20]. The conclusion is that this difference is more quantitative than qualitative.

Jean-Louis Baudry mentions the regressive state as the most remarkable similarity, with the caveat is artificially induced [21]. This state is characterized by suspending disbelief and accepting the play as perception, most of all, provoked by the film viewer's situation mentioned earlier.

Christian Metz concurs with Baudry, but the regression provoked within the viewer by the film does not explain the film viewer's dominant situation, but rather the capacity of the film to organize phantasms through film pictures [22, 23]. Metz mentions investing libidinal energy in film text as the main depth mechanism behind enjoying a film, through the examples of identification, narcissism, voyeurism, and fetishism [22]. Still, when comparing the work of dream and the work of film, Metz emphasizes the second elaboration as the primary force that captivates the viewer in the world of film and allows for the transformation of the incoherent perception of reality coherent fiction [22].

Thierry Kuntzel claims that constructing one film relies upon the same mechanisms as constructing a dream [24]. He takes over Freud's explanation of the work of dream [1] and highlights the primary processes in his explanation of film reception [24].

Given that both film and dream have their own manifested and latent parts analyzing the presentation may lead to film content's unconscious formation [10–14]. When seeking out implicitly, latent film content has become one of the most common reasons for comparing film and dream in film theory.

Analyzing the conclusions brought forth from this comparison, film, dream analogy, and the oneiric metaphors, an incomplete or one-sided comprehension of the dream's dynamic in film theory might become activated, resulting from a similar treatment of dreams in psychoanalysis. Namely, from the very genesis of psychoanalysis, it was possible to recognize dreams' processes and functions without the subconscious's involvement [1]. Yet, there is little attention denoted to its comparison to those related to the subconscious functions and processes. Likewise, they appear neglected in film theory when analogs on the work of dreams come to explain the plot of a film. This why film theorists prioritize sleep as a polygon for unfulfilled wishes and desires rising to the surface from the unconscious [1], while there is neglect in association with other functions and mechanisms of dreams [25]. Testing the assumptions of other theoretical paradigms of psychological sciences, in addition to the psychoanalytic, could lead to significant new insights into the study of film. Note that modern psychological theories on sleep depend on the findings of neuro studies. Film theory can apply to analyze the main findings that indicate the similarities and differences of the neurophysiological bases of dreams and watching movies.

3 Film and Dreams in Neurosciences

It is important to point out that neurological film studies are relatively in infancy that came to exist with the development of functional magnetic resonance (fMRI), which allowed for the recording of the brain's functions of awake subjects over longer intervals [26]. Even so, the functional recording of the nervous systems of film viewers took place for a while, and there is already a significant corpus of new insights into the neurophysiological basis of film reception. It led to the suggestion to establish a new interdisciplinary field of Neurocinematics, the film's neuro study [27]. Conversely, neuro studies of sleep are older, more extensive, and methodologically more grounded than the neuro study of film. It is mostly because during sleep, recording the brain's activity is possible and straightforward to do even using technologically imperfect, while devices of more complexity are necessary for the same purpose during conscious and active viewing of films. Moreover, the film is a concept barely over a hundred years old, while theoretical ideas and deliberations of dreams are significantly older, more extensive, and more influential. Still, the development of the technology needed for recording the brain's morphology and functions in the last several years brings new results that demand new theoretical explanations since they do not fit into the existing clinical theories of dreams [28, 29]. It is where the results of studies relevant to the empirical validation of the similarities between dreaming and film reception occur.

Studies have demonstrated that at present, it is impossible to induce artificial dreams within a subject and that it is challenging to influence the content of a dream using outside stimuli [28], such as watching scary films before going to sleep, tickling, or spraying with water during the REM cycle.

When it comes to dreaming's visual nature, research results suggest that activating the visual cortex during the REM phase of sleep could represent the brain's activity connected with dreaming [30]. Later studies demonstrated that it is incorrect to identify dreams with the REM cycle [29], yet this and similar findings remain strong evidence of dreaming's visual nature. Results also indicated that mental images of objects induce similar activity in the visual cortex and the perception of objects [31]. The same goes for higher-order mental processes that use visual information [32]. Something similar happens with dreams: their visual part, which uses similar principles of coding and direct perception [33], implying that the upcoming period will allow for the possibility of recording, decoding, and observing dreams.

Given the proven visual nature of dreams, what is important for film theory and media in defining the image's concept as the primary concept of analysis is that there is no retinal input [34]. In addition to a lack of information from the retina, there are numerous other neurophysiological similarities between dreaming and visual hallucinations [34].

An important question for both theory and technology of media is that raised by movement during dreaming. Namely, during a dream, the subject can carry out complex movements with an absolute absence of muscular activity. Still, the motor cortex's behavior during such states is neither well known nor researched [35].

Scientists have compared dreaming with mind-wandering moments such as fantasizing, daydreaming, musing, and the like. They found inactivity of similar parts of the cerebrum cortex included in cognitive control and metacognition (medial prefrontal, anterior cingular, and medial temporal lobe). These regions mediate the loss of insight and meta-consciousness, allowing thoughts to take over and disregard stimuli from the objective world [36]. The difference in the prefrontal cortex's activity between wandering minds and dreams is that dreams lead to more profound and more extensive inactivity in the regions responsible for cognitive control and metacognition [36].

There have been attempts to establish the connection between perception and self-representation, one of the primary functions of consciousness, which consists of observing and experience the self in the objective (outer) time and space [37]. Subsequently, scientists found that extensive sensory activity inhibits parts of the brain (primarily parts of the prefrontal cortex) responsible for self-awareness [37].

It is possible to perceive dreams indirectly through evidence from various phenomenology fields. Such an integrational attempt involves developmental, cognitive, neuropsychological, and neurophysiological levels [29, 34]. By which, we can draw the following conclusions. Dreams represent a universal human mental state. The process of dreaming itself is similar to that of imagination (a powerful form of imagination). However, above all visual imagination, dreaming is a higher cognitive function that follows a top-down deductive path of data analysis. Brain activity and metabolism in some sleep stages are very similar to that of a waking state. The primary difference is that there is relatively greater activation of the brain's limbic system and relatively less activity in the cortices, which participate in higher-level cognition. The consciousness of the subject in dreams is almost identical to the consciousness when awake. In dreams, the subject experiences live narrative sensorimotor experiences, while the greatest differences are the absence of self-representation (the subject is not aware that it is in bed and asleep), a lack of will, a rare appearance of certain emotions (such as sadness, remorse, and despondency) on the one hand, and a propensity for part-time hyperemotion and the impossibility of remembering the contents of dreams, on the other hand. The main unsolved question can be singled out as the disconnect from the outside world during dreaming, which should shed light on dreams' insufficiently understood function. Note that dreaming is not a monolithic phenomenon, but rather elaborate that several types of dreams exist. A difference exists between the brain's activity throughout the various periods of sleeping and dreaming, and as such, these fields require further research.

Out of the presented and observed distinctions of dreams thus far, those which might be singled out as attractive to film theory include visual nature, enjoyment, extensive sensory stimulation leading to a loss of self-awareness, that is, a loss of connection with objective time and space. All of the abovementioned suggested that dreaming is indeed similar to watching films. Neurological evidence suggests something similar.

In an experiment in which subjects needed to watch the film *The Good, the Bad and the Ugly* (1966, Sergio Leone) while having their brain activity recorded, there was a great intersubjective similarity (temporal and spatial) of activity in the visual cortex and also specific parts of the brain responsible for more complex information processing [38], e.g., facial recognition. Parts of the brain that did not indicate synchronized activity among the subjects included parts of the prefrontal cortex, the supramarginal and angular gyrus, suggesting different immersion levels among the subjects [38]. Findings of neurostudies on watching films show that, even though the films can take on significant control over our eyes and brain activity, the degree to which it does so varies from film to film and scene to scene. The causes of these variations have yet to be researched [27].

4 Critical Review on the Use of Oneiric Metaphors and Analogies of Film-Dream in Film Theory

The findings mentioned above speak in favor of the justified use of oneiric metaphors and analogies of film-dream as theoretical methods because neurophysiological bases of both processes share a fundamental similarity. Namely, both functionally involve the visual part of the cerebrum's cortex [30, 33, 38], explaining that both appear to be a visual phenomenon in nature. The second similarity is the inhibition of similar parts of the prefrontal cortex [36, 38], which explains the immersion of subjects in the narrative, that is, a loss of self-representation. The difference in the intensity of self-representation is qualitative. It places film reception more closely to moments of mind wandering rather than dreams [36], which is in line with assumptions held by Mitry on the reception of film as a para-oneiric state [20]. However, in addition to the observed similarities, significant differences between the two processes can be noticed, which may have been missed, or which film theory has not paid enough attention to.

The first difference between film and sleep could include the subject's presence in an imaginary world of dreams compared with the film viewer's presence in an imaginary film world. Moreover, a living sense of immediate presence, through observation or participation in events, characterizes a dream [29]. It is, however, not the case with films. Although the nature of this difference could be argued as to whether it is predominantly quantitative or qualitative, the feeling of the subject as to whether it can be included in the plot, interact with the characters, as well as the control over their own actions and behavior in an imaginary world can rarely be found among the film viewer. According to this characteristic, plays and theatrical arts [39] appear to be closer to dreams than films. According to this characteristic, particular similar to dreams are video games, whose artistic potential is insufficiently researched in theory and underutilized in practice. From this viewpoint, virtual reality as an integration of all the media presents another step toward bringing the media experience closer to dreams.

Dependence on the dominant visual nature of films and dreams [10, 11] has led to the neglect of other perceptive aspects when talking about using the film-dream analogy and metaphor. First and foremost, it neglects the full kinesiological proprioceptive sense of the body itself-its stance and movement through the imaginary space of dreams [29], tied with the previously observed presence. Watching films is, based on this characteristic, more similar to the states of fade-out of consciousness while falling into sleepy-hypnagogia, when the body's feeling fades away faster than other parts of its consciousness [29, 40], and every reminder of the body itself leads to the film viewer's departure from his imaginary world of film, that is, the awakening of the subject. Here is where it could be observed as a mistake to rely on the similarity of the positions of the dreamer and the film viewer (dark space, isolation, relaxed corporal position, and absence of movement), which are attributable to the characteristic of film to induce a state of regression and innateness within the subject in film discourse [10–12]. Nonetheless, studies on mind wandering indicate that perceptive stimulation is to blame for dissociating the consciousness from objective time and space [36]. This finding explains how the subjects succeed in immersing themselves in the film discourse not only when they watch the film in a dark movie theater on the big screen, but also, for example, while standing in public transport while watching content on the screens of their mobile phones.

It would be wrong to conclude that the isolation of the subject and the disabling movement are wrong paths to take if the desire is to induce dreamlike experiences. Care is necessary when dealing with (and avoid denial) the relaxation of muscles and sensory deprivation that are frequently needed but not essential [29] for the subject to fall into sleep. On the other hand, the film projection with its lights and lively screen changes and high-intensity sound are also their own blow to the senses. As mentioned several times before, on this issue, film reception and the immersion of the subject in film discourse appear to be more similar to mind wandering than to dreaming-a dreamlike experience without crossing over onto the side of wakefulness. Virtual reality and content that can be consumed in it replace the "inferiority" of the film compared with dreams in which the subject, assisted by technical aids, enters an imaginary space and time of virtual reality. However, a feeling of presence remains, as does the possibility of moving the body, which leads to considerably greater enjoyment and immersion in the imaginary world, that is, greater ignoring of the exterior objective world (compared to film reception). This experience appears much closer to the experience of the subject in dreams.

If the justified approximation of the other sensory modalities is accepted in favor of film and dream's dominant visual nature, what remains is the justification for comparing their content, particularly the syntax and grammar of said content. On the one hand, similarity can be observed since film and dreams are the shapes of adapting reality which figure in pictures [12]. The latent and manifest content of the film, similar to dreams, according to Freud [1], is observed by Kuntzel [24]. Lebovici pronounces both grammars identical since pictures in films and dreams do not mean anything explicitly but are only suggested [16]. Herein lies the problem of the subjectivity of choice and construction materials of film and dream. While without great barriers, it is possible to assume both meanings constructed with the same (subjective) apparatus, the very material upon which the meaning occurs is significantly different. Namely, if Freud's thesis on the *remains of the day* represents material for the building of dreams [1], the question arises about which and whose those remains of the day belong to in the film [10]. Despite many observations that suggest similarities of the procedures combining materials and constructions of film and dream [10–14], not much data to support the idea exists. The subjective origin of the material of dreams carries excess significance compared to the film's objective material. This excess significance comes from the personal experience of the subject with the material of dreams. This difference is quantitative rather than qualitative because it interprets his personal experience in film reception in constructing meaning. However, the significance of these differences is evident in the grammar of the two languages. Namely, in constructing the film, because of his communicative nature, the film author is forced to a certain extent to use the universal use of grammatical processes (symbols, for instance) while the materials of dreams can freely come together. While the author can reach for a completely subjective use of film language, this type of message remains incomprehensible to the film viewer. It is irrelevant for further analysis within the framework of this study. With this in mind, it seems that the process of film making is closer to retelling dreams than dreaming. It means that conditionally speaking, one could speak of the same or similar grammar for the languages of both film and dreams. However, note that there exist at least qualitative differences in the degrees of subjectivity and universality of the two grammars. It is where we must recall Freud, who favored secondary elaboration over the primary processes in the process of retelling films [1].

Authorship is the key difference observed when comparing film and dreams, which must be kept in mind when using the two phenomena' comparisons. Mauerhofer had already drawn attention to this difference [17]. While in sleep, the subject is the sole author of the world, characters, and stories; despite being unconscious of it, in the film, he is merely the consumer. While the theories of texts and those of the performing arts relativize this difference and recognize authorship to the reader as well as spectator [39], that is, they put this difference down to the quantitative rather than qualitative, not enough attention is in film theory dedicated to the creative property of the film viewer—the process of giving meaning to film reception. Under this paper's topic, the process of secondary elaboration seems vital, that is, as a concept with whose detailed analysis and research the observed theoretical hole might form. Namely, while in a conscious state, the subjects' senses appear temporally and spatially consistent stimulations which the awake feelings uninterrupted continuity in the subject, in dreams the images, sounds, thoughts, ideas, emotions, and the like alternate at random. However, it is important to notice that the randomness of appearance and changing of the materials of dreams comes later only when the subject recalls a dream or retells it. There is an almost perfect feeling of continuity and harmony; all incredible transitions practically instantaneously interpret, explain, and situate the course of events constructed in dreams [29]. Similarly, in film, as with dreams, the concept of the conditionally random phenomenon (compared with the objective world and time) of ordering pictures and events appears. Although radically inconsistent compared to the real world yet simultaneously radically consistent with dreams, the film is perceived and processed by the same cognitive apparatus. The mechanism whose primary goal is to establish and maintained meaningful consistency that Freud referred to as secondary elaboration [1], and Metz [22] proclaimed as the dominant force of strength in film perception, while in the perception and production of dreams, it is one of its strengths, but not its primary one. It does not deny the validity of the primary process in the film's reception, which has to date been dominantly in the focus of interest for film theory but draws attention toward the need for a deeper understanding of secondary elaboration, which will lead to deeper insights into the work of primary processes.

Validating such theoretical assumptions leads to disruption over the nature of the film's cognitive analysis process. Therefore, cognitive film theory demonstrates that film's reception alternates bottom-down and top–bottom processes [41], which resembles dream studies' findings [29]. However, these claims must be empirically validated by the neurostudies of film viewers.

Alongside all these observed differences between dream and sleep, it would appear as if sleep and psychology studies generally neglect the potential for studying film reception. Namely, since dreams cannot be provoked nor systematically vary and therefore be used as stimuli on multiple participants, using film as a stimulus in empirical studies ends up being a solution because of the already proven similarities between the nature of film and dream.

Finally, the use of analog and metaphor as theoretical methods has not been satisfactory, and results from such analyses are often incomplete, programmatically encumbered, and imaginary [10]. However, if one uses a phenomenon to compare with it to explain another, that phenomenon must already have a particular corpus of knowledge for which this comparison would be useful. This is why when it comes to explaining and understanding films by comparing them with dreams, one should be aware of the fact that they are insufficiently scientifically explained, while the above findings of the neuroscience of sleep are necessary for further analysis. It appears that the concept of sleep in the theory of film is often arbitrarily used [42], while some of the characteristics assigned to it were not empirically validated later on.

5 Implications

The demonstrated similarities and differences between the cognitive processes in dream and film reception, that is, media content consumption, all have their own implications for film and media theory and the more widespread scientific domain [9]. In the genesis of comprehending film characters, a postulated connection must be observed between its objective nature from social and psychological roles to the imagination and construction of subjective worlds [43, 44]. At first glance, the paradoxical designation of the nature of film as an objective denotation of reality in the function of the imagination and the construction of subjective worlds gains progress through the advancement of technology and expressive media and communications devices. It also postulates the connection between objectivity and subjectivity, that is, the realism of the experience as one of the potential means (even conditions) necessary for qualitatively engaging the cognitive apparatus, which Bazin foresaw as the direction of today's research and the development of the virtual reality media.

In studying the ontological nature of the reception of media content, it must be kept in mind that even the simplest form of perception is the construction, which, to a greater or smaller extent, depends on the physical characteristics of the object but is never fully determined by them. The construction of reality's standards depends on experience and learning, particularly early on in life, and the perception

Fig.1 Conversion of complex real-life objects in the digital environment [46]

apparatus's characteristics. However, the accelerated speed of technology enables the appearance of ever more complex construction media content systems. In generating the content, it is ever more successful in mimicking the cognitive process. This process's hypothetical outcome can be the construction of simulated reality [45], which could not be distinguishable from the real world, that is, its comprehension by foreign subjects. The current technological development level (Fig. 1) justifies the assumption that this hypothesis's day of fulfillment is not far away.

Empirical findings, primarily neuro studies on dreaming and mind wandering [29, 34], suggest that consciousness has, conditionally speaking, two extreme modes of functioning about reality, both physical and subjective. The state in which the subject is oriented for an objective time and space until the end and is completely neglecting subjective reality, which can be called an absolute awake or absolute conscience state, represents the first extreme mode. Analog, the second extreme mode, represents the state, which could be called an absolute dream in which the subject finds himself in a full interior time and space and completely neglects external reality. Such extreme states are only theoretically possible, while in practice, they most often come across passing forms, graded according to their orientation toward one or the other mode of reality. Purely passing forms include state of mind wandering (fantasizing, daydreaming, and musing) and hypnagogia. One such form represents the film's reception, that is, of art and media in general. Based on the intensity of the orientation toward the exterior or interior reality and the intensity of the subject's presence and authorship, we can develop a new classification for art and media. However, understanding the human need for letting themselves experience a narrative could lead to a greater understanding of human nature or at least one part of it that concerns the construction and surrender to inner reality and the need for neglecting objective reality which, to this day, remains insufficiently researched and poorly understood.

While film theory and media often prioritize the observer's role, while dichotomy veers toward resisting observation and participation, the human need to find itself at the border of these two opposite poles must be obvious. It represents the need for symbolic participation or the need to immerse oneself in narrative and phantasmatic participation and action [8, 9]. This need could, to some extent, be due to the unconscious and phantasmatic nature of the libido, that is, the phantasmatic satisfaction of the urge whose realization is neither possible nor allowed in the real world, and the impossibility of the urging acts of the personality to differentiate fantasy from reality [1]. The aforementioned practical impossibility of achieving orientation toward exterior time and space and physiological similarity of the state of the "wandering mind" in its conscious state and dreams speak of the need to consume media content must stress that it provokes subjects immersed into the narrative. The results of neuro studies [36] lead to a conclusion that the perception of stimulation is the primary reason behind the separation of the consciousness from objective time and space, that is, inactivity in parts of the cortex of the cerebrum responsible for cognitive control and metacognition. These findings can explain why the audience prefers and more frequently opts for more dynamic media content, such as a moving picture, compared with a static one. Furthermore, dream as the best way to give in to virtual reality could equally be the ideal of film realism, if not more suitable than absolute awake and physically oriented to the time and space state. An active role of the cognitive process and a degree of their autonomy compared with the physical reality in the construction of experiences of the subject's reality can, in addition to dreams, occur among illusions, hallucinations, and mental illnesses.

In his observation of film's nature through The Myth of Total Film, Bazin predicted technological development whose final goal would be similar to the concept of simulated reality [43]. The current level of technology and media development (Fig. 1) actualizes his concepts and predictions. Digitalization has already enabled a certain transcendence of the material carrier of information, while the transcendence of screens also appears inevitable. When considering the post-screen image concept as a technology that surpasses the technological boundaries of screens, caution is necessary to not fall into the reduction trap. This concept would limit to a picture established in a physical space (e.g., a hologram). Even though this research direction is valid, the basic screen that would need to surpass the post-screen image is the retina. The fact that pictures' formation occurs in the retina cannot be the starting point for observing images' nature. The neurostudies [34] demonstrate that which is evident to anybody who has ever dreamed or experienced a hallucination clearly, an image can be formed independently of the retina. In this sense, it is possible to assume that the upcoming advancements in virtual reality media should depend first and foremost on simulating dreams and on studies based on neuroscience. Today's direction toward attempting to simulate a conscious experience faithfully represents a dead end. Still, given that the subject is the nascent media, this field needs many research projects which will lead science one step closer to understanding film, virtual reality, dreams, and general human nature.

> The real primitives of the cinema existing only in the imaginations of a few men of the nineteenth century, are in complete imitation of nature. Every new development added to the cinema must, paradoxically, take it nearer and nearer to its origins. In short, cinema has not yet been invented!. [43]

6 Conclusion

Neurophysiologic findings of dreaming and watching films conclude that analogies and metaphors of film-dream as a theoretical method in film study are justified and empirically grounded, given that an astonishing similarity exists in the neurological bases of the two processes. However, numerous differences surround the two processes' nature, which demands a more expansive interpretation and new theoretical explanations of the film when using these methods.

The primary similarities between the two processes included the heightened activity of the visual cortex, which explains both processes' visual nature, and inhibition of parts of the prefrontal cortex, which explains the loss of self-representation. The primary differences are the subject's presence in an imaginary world and a connected kinesiological proprioceptive sense of the body itself and a degree of authorship. In dreams, the subject is the exclusive creator of the material of sleep, subjective reality, while when watching films, he is the consumer of something that someone has already created. However, more in-depth research is necessary to understand this degree and principle of authorship of the subject in film reception and yield insights into the process of secondary elaboration.

In addition to numerous studies and findings, the dream is also an insufficiently researched phenomenon. This fact must always be kept in mind when using metaphors or analogies as theoretical methods of the characteristic of sleep for defining any other phenomenon. As a result, it is not uncommon for film theory to have attributed properties that are later not empirically valid or for empirically ungrounded conclusions drawn based on the use of comparing film and dream.

While empirical evidence of the neurostudies of dreaming seek new theoretical explanations for films and point to the possibility and need for connecting and checking theoretical assumptions with quantitative indicators, the studies of sleep and psychology should, however, rely more on researching films that enable systematic variation and exposition of stimuli to a greater number of participants.

The theoretical concepts and empirical evidence introduced lead to the conclusion that the human mind could have two extreme modes of functioning in relation to reality. While the conscious state in which the subject is completely oriented toward the exterior time and space and completely neglects inner reality would be extreme, an absolute dream would represent the other extreme. Of course, these extremes are merely theoretically possible. In practice, they most often run across passing forms that, depending on the degree to which, show orientation toward one or the other mode of reality. Fantasizing, daydreaming, musing or hypnagogia represent examples of passing forms. Film reception, like other art and media content, is another of the passing forms. While through the development of technology, an attempt is to achieve the most faithful simulation of the waking experience, dreams prove that it is possible to feel movement without muscular activity or to perceive an image without retinal activity. A better understanding of dreams and research of their occurrence mechanisms is one of the possible directions of research that could overcome the technological limitations that scientists and engineers face today in achieving the technological development of virtual reality media to simulated reality post-screen images.

Core Messages

- The use of analogies and metaphors of film-dream as a theoretical method in film study appears to be justifiable over an empirically formed ground.
- There is a need for further study of dreaming and film reception, both in film theory and neuroscience.
- The theoretical concepts and empirical evidence conclude that the human mind has two extreme modes of functioning in relation to reality: an absolute conscious and an absolute dream.
- A better understanding of dreams could lead to overcoming the current technological limitations in achieving the development of virtual reality media to the level of simulated reality and post-screen image.
- Integration of distant scientific disciplines such as film theory and neuroscience is possible and offers great hope for both disciplines.

References

1. Freud S (1900) The interpretation of dreams. Franz Deuticke, Vienna
2. Page R, Thomas B (2011) New narratives, stories and storyteling in the digital age. University of Nebraska Press, Lincoln and London
3. Milošević M, Ristić I (2016) New methodology approach to creativity studies. Medias Res časopis filozofije medija 5(8):1237–1250
4. Redfern N (2014) Quantitative methods and the study of film. Research into film. http://nickredfern.files.wordpress.com/2014/05/nick-redfern-quantitative-methods-and-the-studyof-film.pdf. Accessed 01 March 2020
5. Milošević M, Antonijević V (2020) Application of statistical methods in film, media and dramatic text analysis. Kultura 166:302–315
6. Slingerland E (2008) What science offers the humanities: integrating body and culture. Cambridge University Press, New York
7. Manovich L (2015) The science of culture? Social computing, digital humanities, and cultural analytics. http://manovich.net/index.php/projects/cultural-analytics-social-computing. Accessed 01 March 2020
8. Milošević M (2017) Film and dream. Zbornik radova fakulteta dramskih umetnosti 32:27–42
9. Milošević M (2019) Srerbian film in age of new media (2010–2016). Faculty of Aarmatic Art, Universiti of Arts in Belgrade, Disertsation
10. Altman R (1985) Psychoanalysis and cinema: the imaginary discourse. In: Nichols B (ed) Movies and methods. University of California Press, Los Angeles, pp 517–531
11. Casseti F (2001) Film psychoanalysis. Filmske sveske 2–3:19–39
12. Daković N (2001) Dream screen. Filmske sveske 2–3:5–16
13. Rascaroli L (2002) Like a dream. a critical history of the oneiric metaphor in film theory. Kinema 18:5–22
14. Bordwell D (1989) Making meaning: inference and rhetoric in the interpretation of cinema. Harvard University Press, London
15. Glimcher PW (2003) Decisions, uncertainty, and the brain the science of neuroeconomics. A Bradford Book, London
16. Lebovici S (1948) Psychoanalysis and cinema. Revue internationale de filmologie 2(5):49–55
17. Mauerhofer H (1949) Psychology of the film experience. Penguin Film Rev. 8:103–109
18. Langer S (1953) Feeling and form. Scribner, New York
19. Morin E (1956) Cinema or the imaginary man: anthropology of sociological essays. Les editions de Minuit, Paris
20. Mitry J (1963) Aesthetics and psychology of cinema, structures, vol 1. Éditions Universitaires, Paris
21. Baudry JL (1975) The device: metapsychological approaches to the impression of reality. Communications 23:56–72
22. Metz C (1975) The fiction film and its viewer. Communications 23:108–135
23. Metz C (1977) The imaginary signifier. Psychoanalysis and cinema. Union Générale d'Editions, Paris
24. Kuntzel T (1975) The work of the film, 2. Communications 23:136–189
25. Stickgold R, Hobson A, Fosse R, Fosse M (2001) Sleep, learning, and dreams: off-line memory reprocessing. Science 294(5544):1052–1057
26. Kuchinke L, Kappelhoff H, Koelscha S (2013) Emotion and music in narrative films: a neuroscientific perspective. In: Tan S, Cohen A, Lipscomb S, Kendall R (eds) The psychology of music in multimedia. Oxford University Press, London, pp 89–118
27. Hasson U, Landesman O, Knappmeyer B, Vallines I, Rubin N, Heeger D (2008) Neurocinematics: the neuroscience of film. Projections 2:1–26
28. Domhoff W (2002) The scientific study of dreams. APA Press, Washington
29. Nir Y, Tononi G (2010) Dreaming and the brain: from phenomenology to neurophysiology. Trends Cogn Sci 14(2):88

30. Igawa M, Atsumi Y, Takahashi K, Shinichi S, Hirasawa H, Yamamoto R, Maki A, Yamashita Y, Koizumi H (2001) Activation of visual cortex in REM sleep measured by 24-channel NIRS imaging. Psychiatry Clin Neurosci 55:187–188
31. Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci 29(5):1565–1572
32. Stokes M, Saraiva A, Rohenkohl G, Nobre A (2011) Imagery for shapes activates position-invariant representations in human visual cortex. Neuroimage 56(3):1540–1545
33. Horikawa T, Tamaki M, Miyawaki M, Kamitani Y (2013) Neural decoding of visual imagery during sleep. Science 340(6132):639–642
34. Pace-Schott FE (2017) The neurobiology of dreaming. In: Kryger M, TRoth T, Dement CW (eds) Principles and practice of sleep medicine. Elsevir, Amsterdam, pp 563–574
35. De Carli F, Proserpio P, Morrone E, Sartori I, Ferrara M, Gibbs SA, De Gennaro L, Lo Russo G, Nobili L (2016) Activation of the motor cortex during phasic rapid eye movement sleep. Ann Neurol 79(2):326–330
36. Fox K, Nijeboer S, Solomonova E, Domhoff W, Christoff K (2013) Dreaming as mind wandering: evidence from functional neuroimaging and firstperson content reports. Front Hum Neurosci 7:412
37. Goldberg I, Harel M, Malach R (2006) When the brain loses its self: prefrontal inactivation during sensorimotor processing. Neuron 50(2):329–339
38. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640
39. Fischer-Lichte E (2014) The Routledge introduction to theatre and performance studies. Routledge, London
40. Mavromatis A (1987) Hypnagogia: the Unique State of consciousness between wakefulness and sleep. Routledge, London
41. Bordwell D (1989) A case for cognitivism. Iris 9:11–40
42. Curry R (1974) Films and dreams. J Aesthet Art Critic 33(1):83–89
43. Bazin A (1967) What is film?, vol I. University of California press, London
44. Bazin A (1971) What is film?, vol II. University of California press, London
45. Moravec H (1998) Simulation, consciousness, existence, field robotics center at the carnegie mellon robotics institute. http://www.frc.ri.cmu.edu/~hpm/project.archive/general.articles/1998/SimConEx.98.html. Accessed 01 March 2020
46. Sujić B (2018) A look beyond the foreseeable. Faculty of Aarmatic Art, Universiti of Arts in Belgrade, VR content

Miloš Milošević was born in Belgrade, 1980. He has a Master in Psychology and a Ph.D. in Science of Arts. Since 2020, he works as an Assistant professor at Singidunum University in Belgrade in psychology. He has directed over 40 short and documentary films and authored 35 publications in social sciences, humanities, and medical sciences.

Branko Sujić was born in Belgrade, 1980. He is attending D.A. studies of drama and audio-visual arts at the Faculty of Dramatic Arts. Since 2018, he works as an associate professor at the Faculty of Dramatic Arts, artistic field, cinematography, photography, and virtual reality. Branko has authored 22 art projects—films, exhibitions, and virtual reality installations.

Systems View in Engineering Research

6

Aarne Mämmelä and Jukka Riekki

"There is at least a reasonable hope that human evaluations will change toward a survival pattern and that this will happen in the world as a total system".

Kenneth E. Boulding

Summary

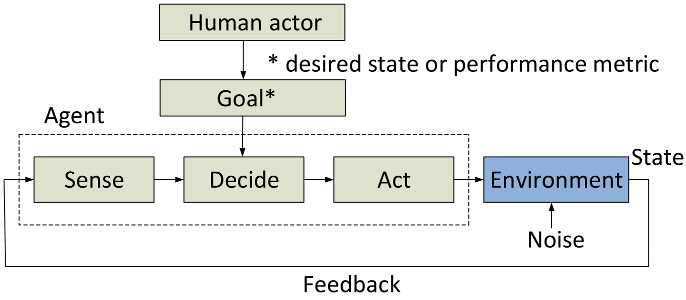
During the last 50 years, the world population has increased by over 110%, when sustainable development would have permitted only a 25% increase. This is clear evidence that global problems are complex, and we need a holistic system view to solve them. We present a tutorial survey on system view in engineering research to facilitate using these ideas in other disciplines, specifically in the topical efforts to build systems that tackle sustainable development challenges. A system is a set of parts and their relationships. The core idea is to build smart systems that optimize resource usage equitably. We present the incommensurable basic resources used by a system and methods to measure its efficiency. We also present the general principles used in smart systems to achieve efficient operation, including control, feedback, optimization and decision making, hierarchy, and degree of centralization. We identify the tragedy of the commons as the main obstacle for sustainable development as it causes overuse of common and free resources. We discuss some open system-level problems and present ideas on applying systems thinking to solve

A. Mämmelä (✉)
VTT Technical Research Centre of Finland, PO Box 1100, 90571 Oulu, Finland
e-mail: aarne.mammela@vtt.fi

J. Riekki
University of Oulu, Oulu, Finland
e-mail: jukka.riekki@oulu.fi

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_6

complex problems. Our major conclusion is the need to use subsidiarity combining weak centralized, hierarchical control and relative autonomy to solve the tragedy of the commons.



The idea of closed-loop (negative feedback) control as a core of smart systems. The goal is given by a human actor.

The code of this chapter is 01010011 01111001 01110011 01110100 01100101 01101101 01110011.

1 Introduction

The Earth forms a closed system in terms of materials and together with the Sun in terms of energy. To have some perspective, the annual energy consumption of the whole world is 0.01% of the annual energy offered by the Sun toward the Earth [1]. The world population is now increasing by 1.1% per year, thus doubling in 64 years [2]. The energy consumption is increasing even faster, 2.3% per year. Part of the used energy is fission energy produced in and on the Earth. The production of energy is causing pollution in the form of radioactive matter and greenhouse gases in the atmosphere. Moreover, insufficient recycling and an inadequate share of renewable materials lead to a lack of scarce resources such as some important metals and cause pollution such as microplastics in oceans.

Sustainability has been regarded as important, at least since the first report of the Club of Rome in 1972 [3]. Sustainability includes ideas such as the end of exponential population growth, the sufficiency of resources, and control of pollution. An example of the need for sustainability can be seen in the Earth Overshoot Day that marks "*the date when humanity's demand for ecological resources and services in a given year exceeds what Earth can regenerate in that year*" [4]. The day has been estimated since 1969. At that time, the day was at the end of the calendar year, and the world population was about 3.6 billion people to whom sustainable development was available. The Earth Overshoot Day is now at the beginning of August, and the world population is about 7.6 billion people. During the last 50 years, the world population has increased by over 110% (1.5% per year) when sustainable development would have permitted only a 25% increase (0.4% per year).

We can estimate the world population to be 70% too high for sustainable development, and the situation is getting worse at an exponential rate, although the annual increase in the world population has reduced to 1.1% per year. Eventually, we must solve the *tragedy of the commons*; otherwise, the whole human population will be in danger when common resources are overused [3, 5, 6]. Hardin devised the term tragedy of the commons for the situation in which actors compete for common and free resources, and everyone is willing to reap the benefits, but no one is ready to pay the costs, usually pollution. The concept has been known since 1833. This is a truly challenging task as it would require partially limiting the independence of the countries with an international social contract to avoid injustice in the use of natural resources and to limit pollution. The reasoning is similar to that of people giving part of their freedom to their home state government in the form of a national *social contract* [7]. People obtain security with this contract as it enables the countries to perform their special task to maintain laws that protect the citizens from injustice [8].

The United Nations has selected sustainability in its long-term agenda [9]. We formulate a vision for a smart and sustainable world by combining sustainable development with the smart world vision [10] that has its origins in the ubiquitous computing vision [11]. Our vision is that prosperity for the people and the planet is achieved with smart and sustainable systems built from sensors, decision units, and actuators. Optimal and equitable use of common resources is emphasized. This vision is empowered by advances in information and communication technologies (ICT), for example, miniaturization, wireless technologies, energy harvesting, Internet of Things and artificial intelligence (AI) software, and in the disciplines closely related to smart systems: computing [12, 13], communications [14, 15], and control [16, 17].

To solve global problems and avoid the tragedy of the commons, resources must be used efficiently, implying smart or intelligent use of basic resources [18, 19]. It is obvious that we need holistic systems thinking in addition to conventional reductive or analytical thinking. A systems view is an abstract high-level view of global problems, whereas the analytical view is a low-level view of the same problems. Analytical methods include deterministic deductive and statistical methods and have many limitations. They cannot be used in the formation of hypotheses, which imply the use of induction or abduction or some form of strong inference using many competing hypotheses [20]. Moreover, the deduction does not produce anything new.

These shortcomings can be overcome by combining the reductive and systems views introducing generalizations, abstract conceptual models, analogies, and common definitions as central tools to provide transferable knowledge useful in new situations and different disciplines, thus improving our creativity [21]. Definitions of useful terms can be found from vocabularies such as [22]. The systems view does not replace the analytical view. A systems view is more general than a reductive view, and thus it can cover more complex phenomena than the reductive view. By a *complex system,* we mean a system made up of a large number of parts that have many, often nonlinear interactions. An example of a systems view is

Mendeleev's periodic table of chemical elements, which needed much earlier reductive work and is clearly a generalization. Generalists and specialists need each other. Generalists using systems view can produce visions and common goals, which help the cooperation of specialists using an analytical view [23, 24].

The central concept in systems view, *system,* can be defined as a set of parts and their causal relationships (Fig. 1) [25]. A system has a boundary with its environment defined by the observer. A system has some function, purpose, or goal in its environment. Example systems include a car, a computer, a software agent, a robot, and a communication link and network. The environment is often called a plant in control theory [17, 26]. When a car's driver is considered as a system, the plant is the car and its environment. The relationships between the parts form the system structure, sometimes called the architecture. The set of properties or features form the system state, for example, the speed of a car. The successive states form the system's behavior. The relationship between a system's input and output forms the system function. A successful system is *stable*, *scalable*, and *efficient*. A stable system has finite outputs for finite inputs; a scalable system can meet greater needs in the future; an efficient system produces good performance with a minimum amount of resources. The general principles used in systems view include control, optimization and decision making, hierarchy, and degree of centralization. The degree of centralization is closely related to the concept of subsidiarity that we suggest as a general solution to the tragedy of the commons. Many of these principles are related to abstract conceptual models.

Linear systems are desirable [27, 28] but not always possible to use when the problems to be solved are not trivial. Linear systems have the following additive property: Two input signals added at the input produce an output that equals the sum of the outputs produced by the same input signals separately. In nonlinear systems, this is not the case, and thus their analysis is much more complicated if not impossible. The fundamental system properties are functionality, performance, dependability, security, privacy, and monetary cost [29]. We focus on system functionality and performance.

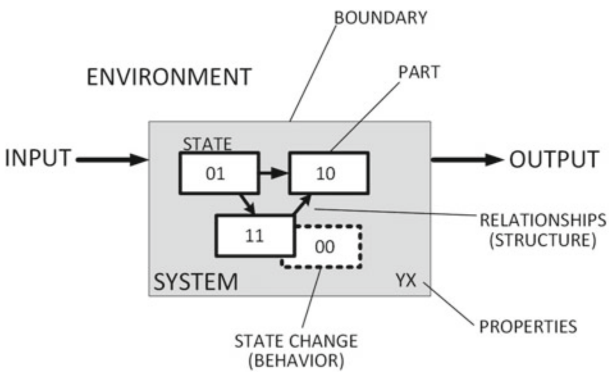


Fig. 1 Definition of a system

Table 1 Scientific societies related to the systems view

Founded	Society
1931	International Council for Science (ICSU), originally International Council of Scientific Unions
1952	International Social Science Council (ISSC)
1956	International Society for the Systems Sciences (ISSS), originally Society for the Advancement of General Systems Theory
1979	Association for Interdisciplinary Studies (AIS)
1990	International Network of Inter- and Transdisciplinarity (INIT)
2018	International Science Council (ISC), merger of ICSU and ISSC

The major scientific societies related to the systems view are listed in Table 1. Paradoxically, systems thinkers do not always know about each other. The ideas presented here have many applications in social systems [30, 31], although they are, unfortunately, not widely known. One of the well-known researchers in this field, Richardson wrote in the preface of his book: *"The seeds of this work were planted almost twenty years ago when I first became aware of the feedback concept. I wondered then and still ponder, how my education could have missed it—the concept was never mentioned in my undergraduate years"* [30].

The chapter is organized as follows. We first present a brief history of systems views using a timeline. Next, we present a description of the basic resources and system principles that include control methods, optimization and decision making, hierarchy, and degree of centralization. Also, we describe the emergence concept and the tragedy of the commons and discuss solutions for the latter, including subsidiarity. We discuss some open system-level problems and present ideas on applying systems thinking to solve complex problems such as those potentially leading to the tragedy of the commons. We end the chapter with conclusions.

2 History of Systems View and Inter- and Transdisciplinarity

The combination of reductive and systems views is our research paradigm. History and inter- and transdisciplinarity are essential parts of the systems view [32, 33], representing a philosophy of engineering. If we know history, we are better prepared to understand the state of art and trends, and therefore we can form a vision for the future, usually for a maximum of ten years. Systems thinking originated in ancient Greece, but modern systems thinking started during the last century, culminating in the foundation of the ISSS society (1956) (Table 1).

Some philosophers and historians of science can be seen as system thinkers [19]. William Whewell (1794–1866) and George Sarton (1884–1956) are examples of philosophers and historians of science whose aim was to form a holistic view of all sciences. Sarton is called the founder of the discipline of the history of science.

His goal was to achieve an integrated philosophy of science that provided a connection between the natural sciences and the humanities, which he called *"the new humanism."* Modern systems thinking was started by researchers such as Alexander Bogdanov (1873–1928) and Ludwig von Bertalanffy (1901–1972). Bogdanov was the pseudonym of Alexandr Malinovsky. His pioneering work has been widely forgotten.

Many of the terms are explained later in this chapter. A timeline of the systems thinking is presented in Fig. 2, which presents independent overlapping efforts, many of which have a long history since their progress has often required improvements in implementation technology. In such cases, the maturation of technology has raised the interest to continue the halted research. The role of interdisciplinarity and transdisciplinary has been strengthened since the 1980s after interdisciplinary was recognized as necessary in solving complex problems [33–35]. Transdisciplinarity is used to support the idea of the unity of science. The foundation of the AIS (1979) and the INIT (2010) shows the raised interest in inter- and transdisciplinarity (Table 1).

Wiener (1948) developed the idea of cybernetics that covered dynamic or time-variant phenomena and combined the results of communication and control. At that time, computers were a new concept, and the idea of computing was almost unknown. AI (1956) was developed to include computing and to distinguish the discipline from cybernetics. However, traditionally AI research has not addressed dynamic phenomena that have been considered in control theory [36]. Robotics combines communication, control, and computing [37]. Boulding developed a hierarchical evolutionary system view by treating the world as a total system [38], which is, at the same time, an example of transdisciplinarity. Now brain science, cognitive science, and AI have been combined into a discipline called intelligence sciences [39].

Typically, the terminology is not unified between disciplines or even within disciplines. For example, robot cars that we now know as self-driving or driverless cars were earlier known as autonomous and smart cars [40, 41]. New terms are developed for the same concepts either due to the lack of knowledge of history or the need to claim the old concepts' novelty. General trends can be seen with Google

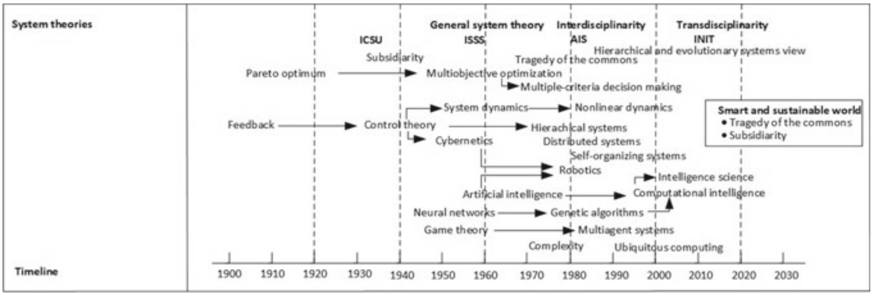


Fig. 2 Timeline of systems view toward a smart and sustainable world

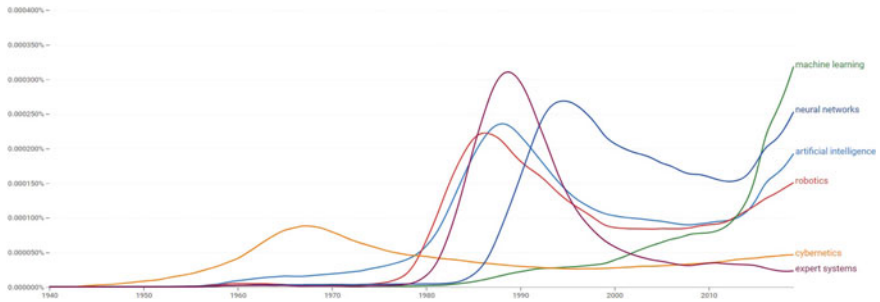


Fig. 3 Popularity of some terms related to smart systems over time (Google Books Ngram Viewer, between 1940 and 2019 from the corpus English 2019 with smoothing of 3)

Books Ngram Viewer (Fig. 3). The popularity of many terms has a peak after their inception, after which interest is reduced and perhaps increased later again. Cybernetics had its peak in 1967, robotics in 1987, AI in 1988 (expert systems dominating at that time), and neural networks in 1994. Machine learning has so far increased its popularity, at least until 2019.

3 Basic Resources

There are six *basic resources*: materials, energy, information (data and control), time (delay), frequency (bandwidth), and space (size) (Fig. 4) [19]. The use of these resources must be optimized, but unfortunately, optimization usually includes conflicting and incommensurable goals and does not guarantee equity in using the resources. For example, energy use should be minimized, but decreasing delays requires more energy. No objective way exists to make a compromise, and thus the monetary cost must be used. However, it is not a scientific problem to tell whether for example, 1 J of energy is more expensive than 1 Hz of bandwidth. In general, incommensurability can be solved only through *survivability*: the systems that fit best to the environment will survive [8, 38]. An example of optimization in our society is the *law of supply and demand* that provides an "invisible hand" that keeps the prices at a reasonable level using an evolutionary approach. Systems must be effective (to do the right things) and efficient (to do the things right), but effectiveness is much more difficult to measure than efficiency [42].

Engineering design starts from *requirements* that form a translation of user needs to technical terms such as functionality and performance. Usually, performance is expressed as *efficiency,* which is the ratio of benefits (e.g., transmitted data bits in communications) to costs (usually the use of basic resources) (Fig. 5) [25]. An example is energy efficiency, which in communications is the number of transmitted data bits per energy unit (in bit/J). An alternative is to express performance directly in terms of the costs, such as a delay (in ms). Information includes data that

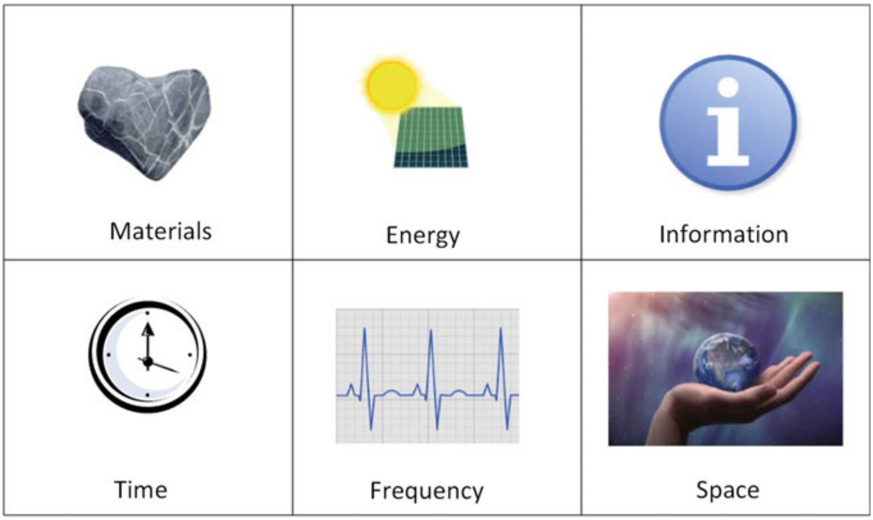


Fig. 4 Basic resources in engineering

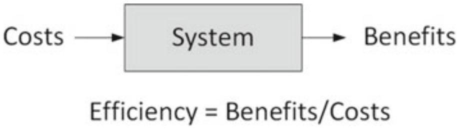


Fig. 5 Definition of efficiency. Costs usually mean the use of basic resources

are benefits in communications and control that is a cost in communications. When requirements are specified, one must consider the fundamental limits of nature, such as the speed of light as constraints [19, 23, 43, 44]. Properties such as cybersecurity and privacy [29] need to be considered, as well.

4 System Principles

The history of smart systems and their structures or architectures is presented in [36, 45]. A set of general system principles or functionalities appear in most systems (Fig. 6) [19]. Commonly used system structures producing characteristic behaviors are called *system archetypes* [3, 46] that are systems that have shown to be useful in various cases and thus survived the test of time in an evolutionary process. They may have both system traps and opportunities. *System traps* are responsible for some potentially dangerous problems that can be transformed, with systems understanding, to produce desirable behaviors [3]. Often, they include positive feedback that must be stabilized with an outer negative feedback loop. Good system

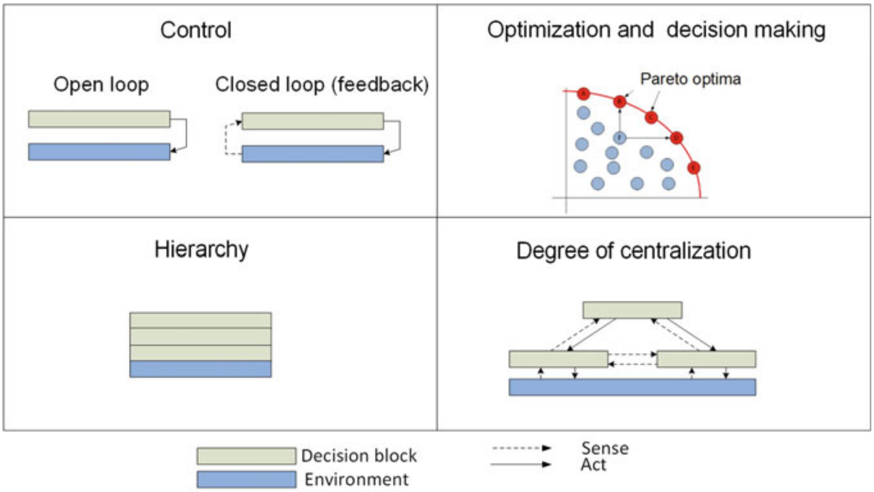


Fig. 6 System principles leading to smart systems

models can be collected in a "system zoo" [47]. The system principles include control [17, 48], optimization and decision making [26, 49, 50], hierarchy [38, 51, 52], and degree of centrality [53, 54]. A smart system usually needs an externally given *goal*, which may be in the form of a set-point value, a reference signal (in adaptive systems), a reference trajectory (in moving vehicles), or improved performance [36, 55]. Complex systems are often called *large-scale systems* [52] or *systems of systems* [56].

The literature on system principles is somewhat disconnected, and the material is in several books, often not referencing each other. Thus, it is difficult to get an overview. Paradoxically even many system theoretical books do not know about each other. Control theory books focus on negative feedback and dynamic systems but ignore other system principles. Positive feedback is usually not discussed in detail in those books. AI does not cover dynamic systems. Feedback does not receive much attention in some machine learning books [57], although it is in the core of learning systems [31, 58]. An alternative to conventional AI is computational intelligence that is often using pattern recognition, but the term pattern recognition is not used systematically in computational intelligence [50, 59]. Social sciences could benefit from an increased understanding of system principles; hence, literature presenting an overview and guiding publications that give more details would be valuable.

4.1 Control

Control can be classified into open-loop and closed-loop (i.e., feedback) control [17]. Feedback is at the core of many smart systems [31, 58]. In general, a control system consists of three blocks: sense, decision, and act (Graphical Abstract). These blocks are connected through the environment, which in control theory is called the plant, the system, or the process to be controlled. Noise represents any uncertainties in the environment. An open-loop control system has no sense block and cannot be very accurate; therefore, feedback control is commonly used. A human decision-maker is called an *actor* in social sciences. The system consisting of the sense, decision, and act blocks is called an *agent* in computer science and a *robot* in robotics. A set of agents is called a *multiagent system*.

The sense block includes the sensors that sense the state of the environment. The block reports the observations to the decision block. The state of the environment is measured by different properties such as temperature, direction, speed, and fuel level of a car. The decision block is called the *controller* in control theory and the *planning* block in robotics. This block implements optimization and decision making. The act block includes actuators that change the state of the environment based on the decisions.

A thermostat controlling a heater is a simple example of feedback control. The targeted temperature given by a human being is the externally given goal; in this case, a set-point value. If the environment's temperature is smaller than the set-point value, the thermostat turns the heater on; otherwise, the heater is turned off. This is called *negative, balancing, correcting,* or *goal-seeking feedback* since the thermostat tries to reduce the difference between the set-point value and the environment's temperature [3, 30, 31]. A similar feedback system is used in cruise control in a car where the set-point value is the desired speed. Negative feedback is known to be stable unless there are long delays in the loop.

Negative feedback creates exponential decay, whereas *positive* or *reinforcing feedback* creates exponential growth and is, therefore, unstable [3, 30, 31]. It is in effect in the case of compound interest. Typical examples of positive feedback are chain reactions such as population and economic growth, bandwagon effect (observable in social media, for example), and pandemic distribution. Such reactions eventually lead to instability unless they are somehow stopped by outer negative feedback. An engineering application of positive feedback is, for example, an oscillator that produces sinusoidal waveforms.

A common example of negative feedback is an election where people give feedback to their representatives through voting. The representatives are decision-makers that produce control actions to the society in the form of laws, punishments, and taxes. Feedback is also used when driving a car, without even noticing it. The sensors are the driver's senses, mainly in the eyes and ears. The actuators are the arms and the legs that control the steering wheel, gearstick, brake, and accelerator (also clutch in older cars). The driver uses feedback to keep the car in the middle of the lane and the speed at a suitable level according to the traffic regulations and the situation. If, for example, the car is drifting toward one edge of the lane, the driver turns the driving wheel to correct the drift, and thus she or he is using negative feedback.

Negative feedback control is usually done iteratively since the desired state can be found directly only in some simple linear systems [36, 60]. The feedback shifts the system toward the desired state or better performance. The basic challenges in using this idea are stability and slow convergence. Therefore, the concept of feedback is not sufficient alone, but hierarchy and different degrees of centralization at different levels of the hierarchy are needed as well. Often the desired state is not available, and the goal is improved performance [55], measured by, for example, energy efficiency, which usually has a definite maximum that acts as a constraint, thus keeping the system stable.

The system model shown in graphical abstract is rather general, and its special cases are shown in Fig. 7. A communication link (e.g., 5G link) offers an example of closed-loop control. Such a link consists of a transmitter, the channel, and a receiver [27, 28, 61]. The propagation channel used for information transmission corresponds to the environment. The transmitter corresponds to the action block, and the receiver corresponds to the sense and decision blocks. The environment may also correspond to the whole network that consists of many links and user terminals. Closed-loop control is used, for example, in transmitter power control, where the power is changed according to the state of the channel.

The open-loop control can be used, for example, in a washing machine where the goal is the wash cycle that does not depend on the dirtiness of the laundry but must be based on earlier experience. The feedback is provided by a human actor, i.e., the washing machine designer, and no feedback is used during the operation. The user may give some instructions or set points using the machine's knobs before the washing begins.

In open- and closed-loop *adaptation or monitoring* no attempt is made to take actions to the environment; thus, the system only observes the environment [62] (Fig. 7). Adaptation-type systems are used, for example, in communication receivers and machine learning. In communications, the goal in adaptation is a reference signal that is ideally a delayed version of the transmitted signal [61]. Open-loop adaptation is sometimes called *one-shot or feedforward adaptation,* and closed-loop adaptation is called *feedback adaptation* [19]. In closed-loop adaptation, the feedback changes the decision block in the form of an adaptive algorithm, not the environment. Feedback is the most common form of adaptation in many applications [62]. In adaptation systems in communications, the channel input is not changed, but the receiver adapts to the channel's state to improve the decisions made in the receiver.

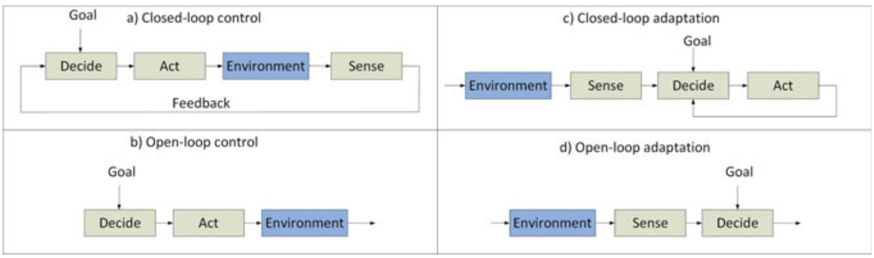


Fig. 7 Comparison of closed- and open-loop control and adaptation

Machine learning (ML) is a subset of AI. In the AI, intelligent agents realize the general closed-loop control model [26], but in the ML, the open- or closed-loop adaptation model is used [57]. In machine learning, the most common learning type is *supervised learning,* where the goal is a *training set* [57]. Usually, learning is the fastest with this form of training. The training set includes examples of correct responses with which the machine learning algorithm makes generalizations to respond correctly to all possible inputs. In *unsupervised learning,* no training set is used, but the decision is based on classifications and statistical analysis of the sensing output, for example, by using pattern recognition. Thus, learning may be slow. *Reinforcement learning* is an intermediate form between supervised and unsupervised learning where no training set is used, but the system is given either positive or negative feedback on the decisions made, corresponding to positive and negative reinforcement, respectively. The algorithm is told whether the decision is right or wrong. This is a form of closed-loop learning. In *evolutionary learning,* the idea of fitness is used. Such solutions will survive that are the best fit for the problem at hand.

4.2 Optimization and Decision Making

The theory of choice is called *decision theory* [63]. *Planning* means that we imagine alternative futures and select the best course of action to achieve a goal state [36, 38]; thus, planning is a combination of optimization and decision making. Technical systems should be optimal to guarantee the efficient use of limited basic resources. However, efficiency in the use of different resources forms mutually conflicting goals for the decision block. The problem is, in general, a complex multi-objective optimization problem [64]. Often several variables have to be optimized in parallel, and the whole problem is nonlinear. The system to be optimized is usually not completely *observable* nor *controllable* [17]. When the system's size increases, the information required for planning can increase exponentially [49, 65]. This is called the *curse of dimensionality* [57]. Several local optima can exist instead of a single global optimum, and thus iterative solutions may converge to a local optimum. Moreover, when a global optimum exists, it may not be unique but a set of optima called *Pareto optima* or *Pareto frontier* (Fig. 6), forming a kind of trade-off beyond which one cannot move. A solution is Pareto optimal if no improvement in any objective can be made without worsening some other objective. In such a case, not all requirements can be fulfilled simultaneously, but the original problem has to be solved with several systems instead of one, each with different performance requirements.

An optimization problem with conflicting goals can be seen as a game where each player has a separate goal. When stability requirements are met, the game converges toward an equilibrium called *Nash equilibrium* that may or may not be Pareto optimal. Nash equilibrium is a situation in a game where no player can gain anything by unilaterally changing her/his own strategy. Both perfect competition and perfect cooperation are unstable situations in social systems, and they tend to drift toward partial cooperation [65]. Cooperation may be limited because of too many players and restricted communication and time [8]. Cybernetics, AI, and game theory do not completely describe human societies, which consist of self-conscious humans that are not always selfish or "rational" as assumed in these theories. An intelligent or *rational agent* is one that can achieve its goals in an uncertain environment [26, 36] with the available resources.

Optimization and decision-making methods include traditional methods, conventional artificial intelligence, and computational intelligence (Table 2). Above them, we have human-level biological intelligence not shown in the Table. For details, see [26, 49, 50, 57]. AI may include conventional artificial intelligence and computational intelligence, but the terminology has not been unified. Hard computing in conventional artificial intelligence means the use of binary logic based on AND, OR, and NOT operations. Sometimes also, statistical methods are used. The opposite of hard computing is soft computing based on pattern recognition used in computational intelligence. Often statistical pattern recognition is not included in computational intelligence. The human brain is good at detecting patterns, more reliable than deductive logic when decisions must be made in an uncertain environment [58]. Generally, no single solution exists for finding the optimum for every problem, but the approach must be selected depending on the application. Exhaustive search as an optimization method is almost always impractical since its complexity depends exponentially on the size of the problem [49, 65]. Even with efficient methods, the optimum is not always possible to find in a finite time.

Different optimization methods represent essentially different learning methods [26, 49, 50, 57, 66]. Nature has developed evolution as a general solution for multi-objective optimization, but this method is extremely slow. Evolution can be simulated using game theory and genetic algorithms, and with them, we can find the Pareto frontier. Since optimization does not produce a unique optimum, a decision must be made using subjective preferences. For example, if a dictator takes all the resources, the situation is still Pareto optimal but also introduces an ethical problem that science cannot solve. In social systems, the preference is usually equity or

Table 2 Optimization and decision-making alternatives in smart systems

Optimization and decision making	Examples	Properties
Computational intelligence	Statistical methods, neural networks, fuzzy systems, evolutionary computing	Pattern recognition including statistical methods, soft computing
Conventional artificial intelligence	Expert systems, case-based reasoning, statistical learning, multiagent systems	Symbol manipulation, deductive logic and statistical methods, hard computing
Traditional methods	Exhaustive search, local search, linear, nonlinear, and dynamic programming, divide and conquer, scalarization, feedback, game theory	

justice, which leads to a system with the highest survivability. Without justice, a revolution may be an unfortunate consequence. In economics, considering subjective preferences leads to welfare economics.

The properties of the environment (Graphical Abstract) form a state that can be interpreted as a pattern. The decision block may hence include three basic subsystems that are analysis (pattern recognition), synthesis (pattern formation), and memory [36, 67, 68]. The memory includes a reduced model of the environment to make learning possible. *Pattern recognition* means automated recognition of regularities in data. In this case, we recognize the state of the environment. The purpose of the decision is to transfer the environment to a state that is better than the earlier one with respect to the goal. This is the *pattern formation* phase in the decision block. Decisions should not be only reactive but also proactive, which is possible when information on the history of the environment is available, and the environment changes slowly enough.

Generally, the speed of learning must be adapted to the rate of changes in the environment. Smart decision making may lead to faster learning. For example, biological evolution is slow, and it does not have any specific direction other than it seems to produce ever more complex structures [38]. Cultural evolution is much faster since it is based on the collective intelligence of human society [38, 58].

4.3 Hierarchies

Hierarchy is a basic method to reduce complexity when the environment has a large number of states, especially when the system has to cope with widely differing ranges in amplitude, time, frequency, or space. For example, in the spatial domain, the range can be the whole planet, different countries, or different municipalities. Hierarchical systems are *modular*, i.e., composed of interchangeable units.

Common system hierarchies can be divided into nested, layered, and dominance hierarchies (Fig. 8) [19, 51]. In the nested hierarchy, the lower-level subsystem is inside the upper-level system. The levels, in this case, are called strata. The nested hierarchy is a description hierarchy having different abstraction levels. This is a common hierarchy in biological systems that consist of atoms, molecules, and cells from the bottom up [69]. Dominance or organizational hierarchy is common in human organizations where the levels are called echelons. An example is a hierarchy of workers, managers, and directors. Originally dominance hierarchy was used in military organizations. Lower levels offer services to the next higher level. Layer hierarchy is a decision hierarchy, and it can be seen as a special case of the dominance hierarchy. The levels are called layers, and each layer dominates only the next lower layer. Such a hierarchical structure is common in some technical systems, such as in robots and communication networks. Systems are often organized into a combination of different hierarchies. An example is matrix hierarchy used in some organizations and in the periodic table of chemical elements.

A hierarchy level should set goals only to the next lower level to avoid deadlock situations [51]. The number of hierarchical levels should be kept at a minimum.

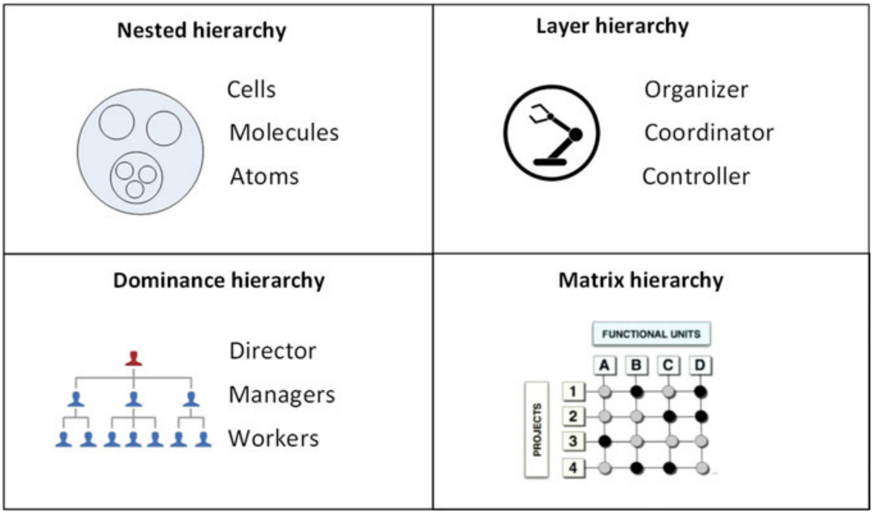


Fig. 8 Common system hierarchies

Usually, three levels are a suitable compromise, but the number can be smaller or larger, depending on the system's complexity. The hierarchy levels usually operate at different ranges and resolutions of the amplitude, time, frequency, and space domains [36, 70]. Ranges are broad and resolutions low at higher levels, whereas ranges are narrow and resolutions high at lower levels [36, 71, 72]. At low hierarchy levels, there may be a need to reduce complexity by decreasing amplitude resolution when processing is made fast [73]. Feedback loops must not be connected (i.e., interfere with each other through the environment) so that instability is avoided. Slower higher levels and faster lower levels improve the stability of hierarchical systems.

Boulding described the world as a total hierarchical system [38]. The levels of this system include physical, biological, social, economic, and political systems. This may also be seen as a hierarchy of corresponding disciplines. Boulding places above these five levels communication and evaluative systems, but we argue that they belong to the social, economic, and political levels where both communications are used, and evaluations and decisions are made. Social systems are based on roles, economic systems on the exchange, and political systems on threats, which is obvious at the international level but also used at the national level in the form of punishments.

In addition to system hierarchies that organize system parts, hierarchies of systems organize systems at different levels based on their characteristics. Two parallel hierarchies of systems are presented with mutual relationships in Fig. 9 [19]. Systems at higher levels are more complex and smarter than lower-level systems [38]. A human actor should always be the highest in the hierarchy since machines understand neither semantics nor context (Graphical Abstract). When a

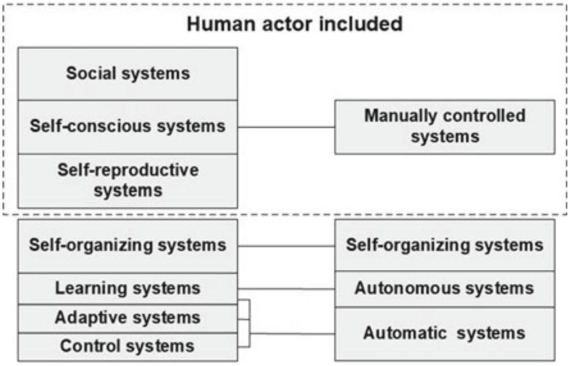


Fig. 9 Two hierarchies of systems using different terminology

group of humans makes the decisions and individual members make more probably right than wrong decisions, i.e., the citizens are politically aware, the probability of the right decisions increases when the group's size increases. Furthermore, a single person as a decision-maker may become incapable of making decisions due to illness or shock, for example. In a Western democracy, common people are the highest in the decision hierarchy since the decision-makers are representatives of the individuals elected by them.

The hierarchy of systems, from the bottom up, is automatic, autonomous, self-organizing, and manually controlled systems. Static and simple dynamic systems such as clockwork below automatic systems are not shown in Fig. 9. Clockworks are simple deterministic automata but usually without feedback. *Automatic systems* do not need any manual intervention, but they may need external control signals. In ML, such systems use supervised learning [57]. *Autonomous systems* are automatic systems that do not need any external control except the goal. *Self-organizing systems* are autonomous systems that can change their structure. *Manually controlled systems* are on top of the hierarchy, and they include a human actor as a decision-maker.

The left-hand hierarchy presents a more detailed division of the lowest levels. Automatic systems can be divided into control systems, adaptive systems, and such learning systems that need an external reference signal using during operation, such as supervised learning. Adaptive systems are advanced control systems that use a reference signal instead of a fixed set-point value (a performance objective, criterion, or metric) and an algorithm in the feedback loop [62, 74, 75]. Learning systems are systems that can change their behavior using earlier experience, and thus memory is mandatory.

All autonomous systems are learning systems. Cognitive systems and artificial intelligence systems are learning systems as well. They do not have true intelligence that would imply self-consciousness. Self-organizing systems are usually distributed [76]. However, the term self-organization implies only that there is no external control, i.e., the whole system is autonomous, although the subsystems may have external control even from a central unit within the system to improve stability. Thus, a self-organizing system can be centralized, and an external goal may improve its stability. Above self-organizing systems, self-reproductive, self-conscious, and social systems are natural systems. Self-production is so far the only known method to produce self-conscious systems (i.e., human beings), and this was achieved through evolution. Human-made systems may form simple social systems based on competition and cooperation but without self-consciousness.

4.4 Degree of Centralization

The degree of centralization may be centralized, decentralized, or distributed (Fig. 10) [47, 53, 54, 77]. In *centralized control*, there is only one agent, including the sense, decision, and act blocks, but in hierarchical, centralized control, there may be agents at each hierarchy level. *Subsidiarity* is a specialized form of centralized, hierarchical control where the lower-level agents have relative autonomy. *Decentralized control* consists of autonomous agents. Since they are competing with each other, such local optimization leads to suboptimal solutions except in the special case of a *free market* without monopolies. In our society, such a free market sometimes requires an intervention of the state government [31]. In *distributed control*, the agents are decentralized, but they also cooperate at least with their nearest neighbors by sharing sensing information and may eventually be able to obtain a global view as in centralized systems. Distributed control has the performance advantage of centralized control but maintains the scalability, ease of implementation, and robustness of decentralized control [77]. Subsidiarity is

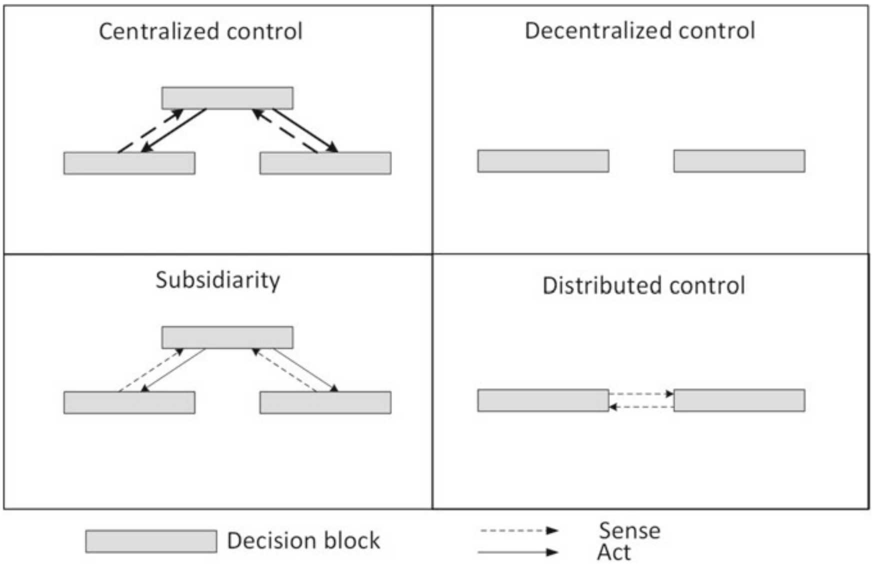


Fig. 10 Different degrees of centralization

another compromise, which benefits from having a centralized control so that the overview is obtained faster than in distributed control.

4.5 Emergence

Emergence is a property of a hierarchical system that cannot be predicted from the subsystems' properties [32]. An example is a temperature that is not a property of individual molecules of gas. One must be careful with emergent phenomena since they may be undesirable and lead to unexpected, chaotic situations. For example, a system with many connected feedback loops may be unstable and behave chaotically. A common solution to this problem is to use a hierarchy with slow upper levels are fast lower levels. In subsidiarity, the subsystems are modular, loosely coupled, and almost isolated so that their mutual interference is minimized. Furthermore, no technical system can be completely autonomous, and human intervention must always be possible.

4.6 Hierarchical Control and Subsidiarity

The tragedy of the commons is serious since it may result in the ruin of all the involved actors because of the resources' overuse. General commons include forests, oceans, and the atmosphere. A historical example of the tragedy is Easter Island, where the culture was destroyed because of the overuse of forests [78]. If a service is equally paid by the users, independently of the actual use, this may lead to the tragedy of the commons. Society has three general solutions to the tragedy of the commons, namely education, privatization, and regulation, and all these are being used to solve the problem [3, 5]. Additional methods are environmental taxes and economic sanctions.

We cannot privatize oceans and the atmosphere, and we should not completely privatize the forests that form a major carbon sink globally. Regulation is a common international solution, for example, in the use of radio frequencies as done by the International Telecommunication Union (ITU). Traffic regulations have also been widely unified. Similarly, international regulation should be used to limit overpopulation, but this would need painful decisions that are also related to social security, which is not well developed everywhere.

Universities have a central role in education and hence in solving the tragedy of the commons. There have been two complementary views, including the German Bildung system having a preference for general education and unity of science and the British liberal education having a preference for specialized professional training [79]. The German system has an important role in developing the systems view that we consider essential in solving global problems. The differences between the two views have been reduced, and they are now seen as complementary elements of Western universities. We emphasize the importance of education as it is a major factor in building the collective intelligence of a country that, in turn, determines the success of the country. The other factors are communication and transportation systems, information storage and retrieval systems, and funding of science and research [58].

Subsidiarity [47, 80, 81] is one general solution to the tragedy of the commons (Fig. 11). According to [82], subsidiarity is "*the principle that a central authority should not be very powerful, and should only control things which cannot be controlled by local organizations.*" The things to be controlled by the highest level include the commons. The idea was invented in ancient times, but the term itself is from the year 1809 in German legal use [83]. This principle has been used in the constitutions of the USA and the EU as well as a reaction against centralization. Subsidiarity seems to be the best and most efficient way to organize a hierarchy and can also be used in technical systems [47]—it is closely related to the degree of centralization of human-made systems. Subsidiarity combines weak hierarchical, centralized control with relatively autonomous lower-level systems. Communication between levels is restricted to the essential [47]. Emergent phenomena are avoided since the feedback loops are decoupled. A good example is a country that is divided into municipalities that have a certain level of autonomy because of the power to collect taxes. Municipalities are responsible for solving local problems near the people where the competence to solve the problems is. These three levels are illustrated in Fig. 11.

Subsidiarity is related to the concept of *bounded rationality* [3]. Individuals usually do not make decisions that add up to the good of the whole, and therefore the upper levels using the subsidiarity principle are necessary. For a complicated problem, hierarchical control decomposes the problem into more manageable units [40, 41, 84]. In a vehicle traffic example, the hierarchy levels have the tasks to assign routes, control traffic flow, coordinate neighboring vehicles, and execute feedback control of the accelerator, braking, and steering [40].

5 Open Problems

The complexity and the slow convergence of learning algorithms is a major open problem of smart systems. This is expected since the complexity and convergence problems are related to the number of states and variables to be controlled in the system and their mutual dependency as in adaptive systems [62, 75].

The second major problem is the complexity of optimization and decision making. The high number of conflicting goals will not make the decision making easy. The stability and convergence rate cannot be easily predicted. There is no straightforward solution to the feedback loop's decision making, especially when the problem is nonlinear and dynamically changing. System dynamics has been studied since the 1600s, but it became popular in the 1950s, and the pioneer was Jay

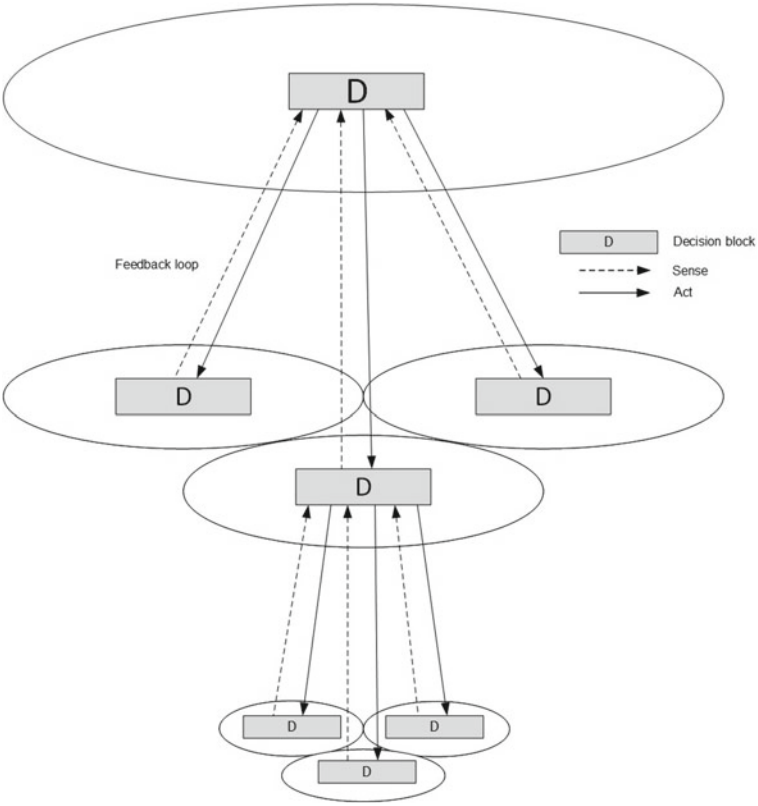


Fig. 11 Subsidiarity to organize the hierarchy toward improved sustainability

Forrester (1918–2016) [85, 86], whose work inspired many others [3, 30, 31, 46, 47]. Nonlinear dynamics have been known since about 1900, but the interest rose again in the 1980s [87].

As mentioned earlier, the complexity of the optimization and decision problem is, in general, exponential, and that is a reasons why hierarchical solutions are popular. In practice, there are too many parameters to control, and communications and time are limited [8]. The implementation technology will no more significantly improve [88], and therefore we cannot expect that any easy solution would be found for the complexity problem. In the end, only the survival will demonstrate the usability of the solutions since the goals are mutually conflicting and incommensurable [8, 38].

A third important problem is the limited observability and controllability of the system [17]. A fourth general problem is the efficient and effective use of limited basic resources, including materials and energy.

6 Solving Complex Problems with Systems Thinking

This chapter was organized around smart systems' principles, including control, optimization and decision making, hierarchy, and degree of centralization. We presented brief historical notes, the basic resources, made some conceptual analysis, and listed for each topic some literature references for further details. The literature is fragmented and uses different terminology for overlapping concepts. Without this kind of tutorial survey, it may be rather difficult to get an overview.

Complex problems require wide expertise and hence interdisciplinarity. The design of complex systems starts from user needs that are translated into system requirements. They describe the functionality, stability, scalability, and performance and some additional properties such as dependability, cybersecurity, and cost. Performance is measured by the efficiency in the use of basic resources, including materials, energy, information, time, frequency, and space. The requirements form a specific research problem that does not exist in natural sciences, which consider natural objects created by evolution. From the requirements, we produce system specifications that form the hypothesis in scientific terminology.

When the research problems and hypotheses are set, the actual research can be conducted by combining reductive and systems thinking. The hypothesis is improved iteratively. Reductive thinking proceeds by reducing a research problem into subproblems, studying these problems with experiments, and refining the hypothesis by generalizing the experiments' results. However, deriving the result by deduction succeeds only in linear and some simple nonlinear systems. The system view is a generalized form of reductive view and provides the tools for deriving complex systems results. The tools include reliable system archetypes and their simulation models that replace the analysis when deduction cannot be made.

Many systems can be classified into a limited number of system archetypes with characteristic behaviors [3, 46]. The archetypes may include both system traps and possibilities. These systems use the general system principles and can be collected to a system zoo that includes simulation models. The system specifications, that is, the hypotheses to solve the research problems can be formed by using the system archetypes. Similarly, a research environment can be developed to support experiments with prototypes. Simulations and experiments replace mathematical analysis in cases where the analysis is not possible.

We recognize the considerable potential in solving some of our greatest global problems, such as overpopulation, with systems thinking. Such problems are system traps generated by positive feedback. Systems thinking and the system principles presented in this chapter provide a basis for solving this type of problem by compensating the positive feedback with negative feedback that works fast enough for avoiding the tragedy of the commons. This would be a promising approach to extend the use of smart systems to support society's decision making. Such systems would observe the society and the environment, propose decisions, and provide feedback on the impact of the decisions realized by society.

7 Conclusion

Major problems in our society are related to overpopulation. The resulting tragedy of the commons can be solved by the subsidiarity principle that is based on weak hierarchical, centralized control and relatively autonomous subsystems. The independence of governments should be partially limited in the same way as people have given part of their freedom to their governments to obtain safety and security in the form of a social contract. We refer to the safety and security of the whole human race, needing an international social contract. We propose systems thinking and smart systems based on the system principles described in this chapter as solutions to global problems and the tragedy of the commons. The researchers need a common vision that is based on a smart and sustainable world, thus forming a research paradigm. A high-level view of systems thinking is needed in solving complex problems. The actual research is carried out reductively; otherwise, the amount of knowledge to master is too large. Our hierarchical approach is based on conceptual analysis using taxonomies and understanding of history and state of the art through annotated bibliographies that include the best books, review papers, and original papers. This approach starts with a vision, proceeds through scenarios and requirements to system specification, and finally, research problems. The decision problem includes multi-objective optimization and decision making with limited basic resources such as materials and energy, and fundamental limits are forming additional constraints. Often the problems are nonlinear and dynamic, but nonlinearity should be avoided as much as possible by system design. We have observed that a general theory of nonlinear multidimensional causal relationships is missing, and therefore the systems must be implemented hierarchically to avoid undesirable emergent chaotic phenomena. Communication between disciplines is insufficient and should hence be increased through interdisciplinary research. Systems thinking, system principles, such as hierarchy and subsidiarity, and smart systems could offer potential solutions to other disciplines than technological ones, including decision support, to solve the global problems of our society.

Core Messages

- Our vision is that we are moving toward a smart and sustainable world where the limited resources are used efficiently.
- Systems thinking and smart systems can help in solving complex problems.
- Applying the general system principles such as hierarchy, feedback, and the subsidiarity principle can help to solve some of the major problems of our time that otherwise can lead to the tragedy of the commons.

Acknowledgements Discussions with Adrian Kotelba and Kari Leppälä are acknowledged. Jukka Riekki was financially supported by the Academy of Finland 6G Flagship grant 318927.

References

1. International Energy Agency (2019) Global energy & CO_2 status report. https://www.iea.org/geco. Accessed 13 July 2020
2. United Nations (2019) World population prospects 2019. https://population.un.org/wpp. Accessed 13 July 2020
3. Meadows DH, Wright D (eds) (2008) Thinking in systems: a primer. Chelsea Green Publishing, White River Junction, VT
4. Global Footprint Network (2020) Earth overshoot day. https://www.overshootday.org. Accessed 13 July 2020
5. Hardin G (1968) The tragedy of the commons. Science 162(3859):1234–1248
6. Hardin G (1998) Extension of 'the tragedy of the commons.' Science 280(5364):682–683
7. Honderich T (ed) (2005) The Oxford companion to philosophy, 2nd edn. Oxford University Press, Oxford, UK
8. Siegfried T (2006) A beautiful math: John Nash, game theory, and the modern quest for a code of nature. Joseph Henry Press, Washington, DC
9. United Nations (2015) Transforming our world: the 2030 agenda for sustainable development. https://www.un.org/sustainabledevelopment/development-agenda. Accessed 13 July 2020
10. Ma J, Yang LT, Apduhan BO, Huang R, Barolli L, Takizawa M (2005) Towards a smart world and ubiquitous intelligence: a walkthrough from smart things to smart hyperspaces and ubickids. Int J Pervasive Comput Commun 1(1):53–68
11. Weiser M (1991) The computer for the 21st century. Sci Am 265(3):94–104
12. Hennessy JL, Patterson DA (2012) Computer architecture: a quantitative approach, 5th edn. Morgan Kauffman, Waltham, MA
13. Turing D (2018) The story of computing: from the abacus to artificial intelligence. Arcturus Publishing, London
14. Huurdeman A (2003) The worldwide history of telecommunications. Wiley, Hoboken, NJ
15. Kurose JF, Ross KD (2013) Computer networking: a top-down approach featuring the internet, 6th edn. Addison-Wesley, Boston, MA
16. Bernstein DS (2002) Feedback control: an invisible thread in the history of technology. IEEE Control Syst Mag 22(2):53–68
17. Ogata K (2010) Modern control engineering, 5th edn. Prentice-Hall, Boston, MA
18. Wilenius M (2014) Leadership in the sixth wave—excursions into the new paradigm of the Kondratieff cycle 2010–2050. Eur J Futures Res 2(36):1–11
19. Mämmelä A, Riekki J, Kotelba A, Anttonen A (2018) Multidisciplinary and historical perspectives for developing intelligent and resource-efficient systems. IEEE Access 6:17464–17499
20. Platt JR (1964) Strong inference. Science 146(3642):347–353
21. Epstein D (2019) Range: why generalists triumph in a specialized world. Macmillan, London
22. Software and Systems Engineering Vocabulary (2020) IEEE Computer Society and ISO/IEC JTC 1/SC7. https://pascal.computer.org. Accessed 13 July 2020
23. Barrow JD (1998) Impossibility: the limits of science and the science of limits. Oxford University Press, Oxford, UK
24. Choi BCK, Pak AWP (2007) Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. Clin Invest Med 30(6):224–232
25. Hubka V, Eder WE (1988) Theory of technical systems: a total concept theory for engineering design. Springer, Berlin
26. Russell S, Norvig P (2010) Artificial intelligence: a modern approach, 3rd edn. Prentice-Hall, Upper Saddle River, NJ
27. Haykin S, Moher M (2009) Communication systems, 5th edn. Wiley, New York
28. Ziemer RH, Tranter WH (2015) Principles of communications: systems, modulation, and noise, 7th edn. Wiley, Hoboken, NJ

29. Avizienis A, Laprie JC, Randell B, Landwehr C (2004) Basic concepts and taxonomy of dependable and secure computing. IEEE Trans Dependable Secure Comput 1(1):11–33
30. Richardson GP (1991) Feedback thought in social science and systems theory. University of Pennsylvia Press, Pennsylvania, PA
31. Sterman JD (2000) Business dynamics: systems thinking and modeling for a complex world. McGraw-Hill, Boston, MA
32. Checkland P (1999) Systems thinking, systems practice, new. Wiley, Chichester, UK
33. Klein JT (1990) Interdisciplinarity: history, theory, and practice. Wayne State University Press Detroit, MI
34. Choi BCK, Pak AWP (2006) Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. Clin Invest Med 29(6):351–364
35. Repko AF, Szostak R (2020) Interdisciplinary research: process and theory, 4th edn. SAGE, Thousand Oaks, CA
36. Albus JS, Meystel AM (2001) Engineering of mind: an introduction to the science of intelligent systems. Wiley, New York
37. Bekey GA (2005) Autonomous robots: from biological inspiration to implementation and control. MIT Press, Cambridge, MA
38. Boulding KE (1985) The world as a total system. SAGE, Beverly Hills, CA
39. Shi Z (2009) On intelligence science. Int J Adv Intell 1(1):39–57
40. Varaiya P (1993) Smart cars on smart roads: problems of control. IEEE Trans Automat Contr 38(2):195–207
41. Badue C, Guidolini R, Carneiro RV, Azevedo O, Cardoso VB, Forechi A, Jesus L, Berriel R, Paixao T, Mutz F, Veronese L, Oliveira-Santos T, De Souza AF (2019) Self-driving cars: a survey. https://arxiv.org/pdf/1901.04407.pdf. Accessed 13 July 2020
42. Zidane YJT, Olsson NOE (2017) Defining project efficiency, effectiveness and efficacy. Int J Managing Projects Bus 10(3):621–641
43. Markov IL (2014) Limits on fundamental limits to computation. Nature 512:147–154
44. Mämmelä A, Anttonen A (2017) Why will computing power need particular attention in future wireless devices? IEEE Circuits Syst Mag 17(1):12–26
45. Kotseruba I, Tsotsos JK (2020) 40 years of cognitive architectures: core cognitive abilities and practical applications. Artif Intell Rev 53:17–94
46. Senge PM (2006) Fifth discipline: the art and practice of the learning organization, rev. Doubleday, New York
47. Bossel H (2007) Systems and models: complexity, dynamics, evolution, sustainability. Books on Demand, Norderstedt, Germany
48. Dorf RC, Bishop RH (2017) Modern control systems, 13th edn. Prentice-Hall, Upper Saddle River, NJ
49. Michalewicz Z, Fogel DB (2004) How to solve it: modern heuristics, 2nd edn. Springer, Berlin
50. Eberhart RC, Shi Y (2007) Computational intelligence: concepts to implementations. Morgan Kauffman, Burlington, MA
51. Mesarovic MD, Macko D, Takahara Y (1970) Theory of hierarchical, multilevel systems. Academic Press, New York
52. Filip FG, Leiviskä K (2009) Large-scale complex systems. In: Nof SY (ed) Springer handbook of automation. Springer, Berlin, pp 619–638
53. Coulouris G, Dollimore J, Kindberg T, Blair G (2012) Distributed systems: concepts and design, 5th edn. Addison-Wesley, Boston, MA
54. Ghosh S (2015) Distributed systems: an algorithmic approach, 2nd edn. CRC Press, New York
55. Boulding KE (1978) Ecodynamics: a new theory of societal evolution. SAGE, Beverly Hills, CA

56. Nielsen CB, Larsen PG, Fitzgerald J, Woodcock J, Peleska J (2015) Systems of systems engineering: basic concepts, model-based techniques, and research directions. ACM Comput Surv 48(2):18:1–18:41
57. Marsland S (2015) Machine learning: an algorithmic perspective, 2nd edn. Chapman and Hall/CRC, New York
58. Stonier T (1992) Beyond information: the natural history of intelligence. Springer, Berlin
59. Friedman M, Kandel A (1999) Introduction to pattern recognition: statistical, structural, neural and fuzzy logic approaches. Imperial College Press, London
60. Makhoul J (1975) Linear prediction: a tutorial review. Proc IEEE 63(4):561–580
61. Proakis JG, Salehi M (2008) Digital communications, 5th edn. McGraw-Hill, New York
62. Widrow B, Stearns AD (1985) Adaptive signal processing. Prentice-Hall, Englewood Cliffs, MA
63. Tsoukias A (2008) From decision theory to decision aiding methodology. Eur J Oper Res 187:138–161
64. Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. Struct Multidiscipl Optim 26(6):369–395
65. Boulding KE (1981) Evolutionary economics. SAGE, Beverly Hills, CA
66. Talbi EG (2009) Metaheuristics: from design to implementation. Wiley, Hoboken, NJ
67. Mendel JM, Fu KS (eds) (1970) Adaptive, learning, and pattern recognition systems: theory and applications. Academic Press, New York
68. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
69. Miller JG (1973) Living systems. Q Rev Biol 48(2):63–91
70. Rexford J, Greenberg A, Hjalmtysson G, Maltz DA, Myers A, Xie G, Zhan J, Zhan H (15–16 Nov 2004) Network-wide decision making: toward a wafer-thin control plane. In: Proceedings of workshop on hot topics in networks (HotNets), San Diego, CA
71. Saridis GN (1979) Toward the realization of intelligent controls. Proc IEEE 67(8):1115–1133
72. Holma H, Toskala A (2011) LTE for UMTS: evolution to the LTE-Advanced, 2nd edn. Wiley, Chichester, UK
73. Wu YS (1989) ‘Constant capacity,’ DSP architecture: an historical perspective. In: Chan YT (ed) Underwater acoustic data processing. Kluwer, Dordrecht, The Netherlands, pp 609–627
74. Claasen TACM, Mecklenbräuker WFG (1985) Adaptive techniques for signal processing in communications. IEEE Commun Mag 23(11):8–19
75. Haykin S (2013) Adaptive filter theory, 5th edn. Prentice-Hall, Englewood Cliffs, NJ
76. Prehofer C, Bettstetter C (2005) Self-organization in communication networks: principles and design paradigms. IEEE Commun Mag 43(7):78–85
77. Frampton KD, Baumann ON, Gardonio P (2010) A comparison of decentralized, distributed, and centralized vibro-acoustic control. J Acoust Soc Amer 128(5):2798–2806
78. Diamond J (2005) Collapse: how societies choose to fail or succeed. Penguin, New York
79. Holford J (2014) The lost honour of the social dimension: Bologna, exports and the idea of the university. Int J Lifelong Educ 33(1):7–25
80. Endo K (1994) The principle of subsidiarity: from Johannes Althusius to Jacques Delors. Hokkaido Law Rev 44(6):563–652
81. Follesdal A (1998) Subsidiarity. J Polit Philos 6(2):190–218
82. Oxford Learner’s Dictionaries (2020) Oxford University Press, Oxford, UK. https://www.oxfordlearnersdictionaries.com. Accessed 13 July 2020
83. Economides K (2012) Centre-periphery tensions in legal theory and practice: can law and lawyers resist urban imperialism? Int J Rural Law Policy 2(Special Edition):1–8
84. Hill DJ, Liu T, Verbic G (22–26 July 2012) Smart grids as distributed learning control. In: Proceedings of the IEEE power and energy society general meeting, San Diego, CA
85. Forrester JW (2007) System dynamics—a personal view of the first fifty years. Syst Dyn Rev 23(2/3):345–358
86. Forrester JW (2007) System dynamics—the next fifty years. Syst Dyn Rev 23(2/3):359–370

87. Perona E (2005) Birth and early history of nonlinear dynamics in economics. Revista de Economia y Estadistica 43(2):29–60
88. Waldrop MM (2016) More than Moore. Nature 530:145–147

Aarne Mämmelä received the degree of D.Sc. (Tech.) (with honors) from the University of Oulu in 1996. He was with the University of Oulu from 1982 to 1993. In 1993, he joined the VTT Technical Research Centre of Finland in Oulu. Since 1996, he has been a Research Professor of wireless communications. He has visited the University of Kaiserslautern in Germany in 1990–1991 and the University of Canterbury in New Zealand in 1996–1997. Since 2004, he has been a Docent (equivalent to Adjunct Professor) at the University of Oulu. In 2014–2018, he was an Editor of the *IEEE Wireless Communications* and, in 2016–2018, a member of the Research Council of Natural Sciences and Engineering in the Academy of Finland. He has given lectures on research methodology at the University of Oulu for about 20 years, including the systems approach. He has been an advisor to ten doctoral students and published over 40 journal papers, 14 book chapters, and almost 100 conference papers.

Jukka Riekki received the D.Sc. (Tech.) degree from the University of Oulu, Finland, in 1999. Since 2001, he has been a professor of software architectures for embedded systems at the University of Oulu and since 2014, the Dean of the Faculty of Information Technology and Electrical Engineering. He visited Intelligent Systems Division, Electrotechnical Laboratory, Tsukuba, Japan, in 1994–1996. In his PhD thesis, he studied the reactive task execution of mobile robots. In 2000, he started to study context-aware and ubiquitous systems. In recent years, he has shifted his research focus to the Internet of Things and edge computing, emphasizing distributed and resource-limited systems. He has been an advisor of ten doctoral theses. He has published 45 journal papers, six book chapters, and over 180 conference papers.

7 Molecular Dynamics Simulations: Concept, Methods, and Applications

Mohammad Sufian Badar, Shazmeen Shamsi, Jawed Ahmed, and Md. Afshar Alam

"Our future world will have to find equilibrium in the technology pendulum swing."

Stephane Nappo.

M. S. Badar (✉)
Department of Bioengineering, University of California, Riverside, CA, USA
e-mail: msbadar@engr.ucr.edu; director_academic@salfiainstitute.edu.in

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Riverside, CA, USA

Salfia Paramedical Institute, Darbhanga, India

S. Shamsi
Department of Computer Science, Jamia Millia Islamia, Jamia Nagar, New Delhi 10025, India
e-mail: shazmeen184002@st.jmi.ac.in

PGD (Big Data Analytics), School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

J. Ahmed
Department of Computer Science, Jamia Hamdard University, Delhi, India
e-mail: jahmed2047@jamiahamdard.ac.in

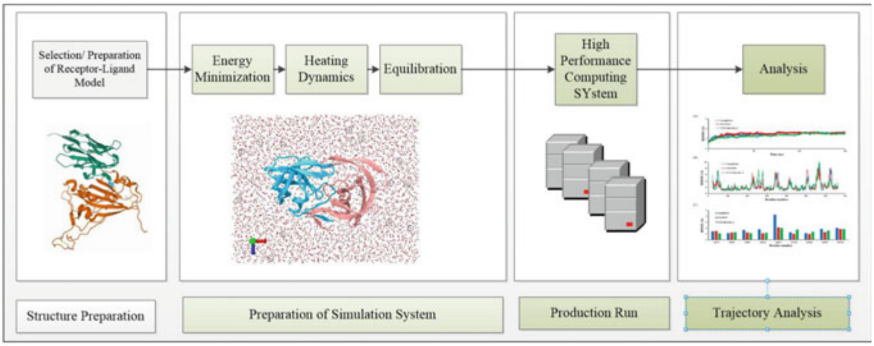
Md.A. Alam
Jamia Hamdard, New Delhi, India
e-mail: aalam@jamiahamdard.ac.in

School of Engineering Science and Technology, Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_7

Summary

Molecular dynamics (MD) is a computer simulation that deals with biological molecules, such as proteins and nucleic acid, and visualizes their movement in atoms and molecules. Computer simulation is executed with these atoms and molecules that are capable of interacting with each other over time and thereby can define the dynamic evolution of the system. MD simulation mimics the changes in biological molecules' structures over a given time, giving us atomic insights into the change in structure. This data helps us understand biological functions. These simulations give us comprehensive information about the fluctuations and flexibility of the proteins and nucleic acids under study. These approaches are applied to thoroughly study the organization and dynamics of biological molecules, their complexes, and conformational changes in proteins and nucleic acids. Many mysteries, on the femtoseconds scale, have been revealed through the study of these conformational changes. These methods are applied in chemical physics, materials science, and biophysics. MD simulations are often used in computational biology to generate a comprehensive understanding of interactions between proteins and their ligands and address how much these interactions are flexible and shape conformational changes in molecules when a particular mutation is introduced. Currently, it is being used to determine the tertiary structure of proteins from x-ray crystallography and NMR (or Nuclear Magnetic Resonance, a technique used in analytical chemistry for determining the structural properties and purity of samples) experiments.



The molecular dynamics simulation process.

The code of this chapter is 01001101 01101111 01101100 01100101 01100011 01110101 01101100 01100001 01110010 00100000 01100100 01111001 01101110 01100001 01101101 01101001 01100011 01110011 00100000.

1 Introduction

Paraphrasing Stephane Nappo, the Global Head of Information Security for Société Générale International Banking pole, this chapter attempts to address how the world of research in biological science is heading toward an incredibly revolutionary technology by amalgamating the focal aspects of Biological Science with Physical, Chemical, and Computer Science which aids it in manifesting the core characteristic of modern interdisciplinary research thus giving life to the technical outlook of Nappo's statement.

The ultimate goal of man is the sheer modernization of lifestyle for his comfort. Science aims at developing tools to fulfill this objective. Everything in the universe is evolving- be it nature or some human-made technology. The only difference is that nature evolves gradually while human-made technology is evolving at a much faster pace. Today the cure, antidotes, or therapies to almost all kinds of diseases are possible. Experimentation is important to gain insights into what the outcome of a particular trial could be. Determination of the structure of biomolecules under different conditions is a requisite to develop a proper understanding of the interactions between them to develop drugs for different diseases as per the need of the situation. It is experimentally impracticable to accurately determine the time-dependent behavior of biological molecules under real-time laboratory conditions, but developing a proper understanding of the dynamic behavior of complex biological processes such as protein folding and stability, conformational changes, ion transport, the central dogma of life, enzymatic reactions, etc. dynamic. is important for the development of drugs, therapies, and techniques that help cure diseases. Simulations help us to steer clear of this problem by using computational techniques. It also helps overcome time and cost issues in the long term. Experimental biological samples are pretty costly, and if a trial goes unsuccessful, the researchers have to bear a heavy loss in terms of time and money. Also, sometimes an experiment needs to be done within a limited period, take, for example, the development of drugs against coronavirus disease (COVID-19). Researchers need to do this task as soon as possible. However, considering the newness of this virus and the fact that not much is known about its reaction with different molecules in the human body, it is frustrating for the researchers to find the exact point of action of the virus experimentally because, one, it will take much time and keeping in mind the number of new cases that are emerging each day, the scientists cannot afford to waste time. Two, it will be very costly and risky to get real-life biological samples, and clinical trials take years to approve the validation of newly designed drugs. This is where simulation techniques come to the rescue. The interaction of the different biological molecules with the designed drug molecules can be seen using computational simulation tools, and the lead compounds can be separated. Then laboratory examination of only the screened compounds could be done further. This method saves time, money, and energy for the researchers, and the fact that further experimentation is done on the screened molecules only, there is a much higher chance of getting positive/desired results.

Simulation of biological molecules was unknown until as late as the 1950s, but within the next ten years from then, it was one of the hottest topics in the research world. The literal meaning of ‘simulation’, as we all know, is the imitation of an anticipated event. *Molecular Dynamic Simulation* is executed with computer techniques to apprehend the dynamicity of biomolecules. These methods visualize atoms and molecules when interacting with each other for a secure duration of time and analyze their physical movement and chemical interactions. Therefore, they help us realize the structure, fluctuations, flexibility, conformational changes, dynamics, and thermodynamics of simple biological molecules as well as their complexes. Understanding these complex biomolecular motions is doubtlessly pertinent to drug discovery [1]. The initial ‘lock-and-key’ mechanism of ligand binding proposed by Emil Fischer in 1890, in which a stationary, fixed receptor was assumed to house a small molecule without going through any conformational rearrangements, has now been forsaken to accept new binding models that consider not only the conformational changes but also the random motions of ligands and receptors [2–6], thus proving Richard Feynman’s statement true. He was a Nobel Prize recipient (1965) in Physics and said, “All things are made of atoms, and that everything that living things do can be understood in terms of the jiggling and wiggling of atoms” [7]. Today, biophysics is a field devoted to comprehending the true essence of this jiggling and wiggling of biological molecules.

1.1 Aim

The goal of molecular dynamic simulation is to predict the behavior of atoms in a biological system and how they move as a time-dependent function, thereby providing the ultimate details concerning the atoms based on algorithms of physics that govern the interatomic interactions [8]. Through this, we hope to discern the properties of molecules concerning their structure and their conduct under different conditions. It serves as an important suffix to the lab experiments, saving time, cost, and labor of the scientists and bridges the gap between the latest technological advancements in the modern scientific community and the conventional experimental scientists. It aims at lowering the amount of guesswork and fittings traditional scientists make and helps them get an idea about the simulations that are difficult or unfeasible in the laboratory. We should always keep in mind that it is possible that one might not necessarily have a flawlessly realistic molecular model. However, the model should be able to portray the essential properties of physics and chemistry and also follow the concerned laws of mathematics along with possessing the correct biological attributes, and that should be enough.

1.2 Brief History

Alder and Wainwright first introduced Molecular Dynamic Simulation in 1957–1959 to understand hard spheres’ interactions through thorough study. Even though the first proper usage of simulation dates back to 1964 when Rahman et al. initiated developing real-world liquid argon. The numerical methods used for this process were developed much before, preceding the use of computers. In 1969, Barker and Watts first performed the Monte Carlo simulation of water, while McCammon et al. in 1977 performed the first MD protein simulation. The protein of interest was the bovine pancreatic trypsin inhibitor (BPTI). Duan and Kollman, in the 1990s, made an amazing revelation by discovering the folding mechanism of villin protein by applying techniques of molecular dynamics simulation, and this achievement is considered a landmark event of this field [9].

Now you must be wondering what Monte Carlo Simulation is? For that, we need to understand that there are two main classes of simulation techniques: the molecular dynamic (MD) simulation and Monte Carlo (MC) simulation. Moreover, there are other composite techniques that integrate the features of both these MD and MC depending upon the need of the research [10]. For a simulation of low-density systems like gas, where the molecules possibly get trapped in low-energy conformations, Monte Carlo simulations are preferable, while MD simulation is the choice technique for the simulation of liquids [11]. Further discussion on MC is beyond the scope of this chapter.

2 Concepts

Computer simulation for studying the dynamic behavior of molecules to comprehend the enigma behind the complexity of the biological world is a demanding task. It necessitates the need for optimally developed models capable of mimicking the cellular environment. These physical forces can simulate the laws of physics and thermodynamics and provide dynamicity to the model and heavy computations keeping in view the temporal aspect of the technique. Today, tools have been developed for molecular modeling, energy calculations, algorithms to simulate the real systems' chemical aspect, docking-scoring techniques, etc., thereby making the whole technique robust. To make the simulation naturalistic, the structure is placed in a "bath" of thousands of water molecules. Let us generate a fundamental idea about this incredibly amazing technology-enhanced technique.

2.1 Molecular Modeling

Molecular modeling is one of the fastest spreading techniques in computational biology, which encompasses all the tasks from visualization, derivation, manipulation, and representation of the structures of molecules keeping in view the physical and chemical properties that depend on these structures. As per recent studies, the modeled molecules should simulate their behavior, taking into account the equations of classical and quantum physics [12]. At present, the total number of entries in the UniProtKB/TrEMBL database is 184,998,855, while in PDB, it is

Table 1 The existence of proteins identified at different levels of information

a	1
b	2
c	3
d	4

(Prepared with data from UniprotKB/TrEMBL database as of July 25, 2020)

166891, which can be seen in the table given below, which has been taken from the UniProtKB/TrEMBL database.

Table 1 represents the number of entries of proteins in UniProtKB/TrEMBL database at different levels of its existence as of July 25, 2020.

MD simulation considers molecules as a ball-on-spring model. This model is apt to simulate the dynamic behavior of the molecules. Molecular modeling helps generate the structures of biomolecules by supplying the geometrical coordinates of biomolecules available as NMR or X-Ray crystallographic structures. However, if the ready-made structures are unavailable, one can easily deduce them by using computational algorithms and then assigning the x-, y- and z- coordinates to the molecules from the knowledge of their geometry. Three prime methods used for modeling are the ab-initio method, threading, and homology modeling.

2.2 Molecular Interaction and Force Field

MD simulation requires equations of motion for classical mechanics, which can in the simplest form be written as

$$m_i r_i = f_i$$

where $f_i = -\bar{\delta} \overset{\delta}{\underset{r_i}{u}}$

For this calculation, we should numerically know f_i as the force that acts on the atoms, which, in turn, results from a potential energy $U(r^N)$, where $r^N = (r_1, r_2, \ldots r_N)$ stands for the entire set of the geometrical 3 N coordinates of each atom.

For this potential energy calculation, we first need to develop a clear concept of its functional form, the force field. Force field can be understood as an empirical set of energy functions that helps us get an understanding of the energy associated with the interaction between atoms [13]. Typically, a force field is the summation of bonded and non-bonded terms or covalent and non- covalent interactions among the atoms and molecules as,

$$E_{\text{Total}} = E_{\text{Stretch}} + E_{\text{Bend}} + E_{\text{Torsion}} + E_{\text{Electrostatic}} + E_{\text{van der Walls}} + E_{\text{Hydrogen Bond}}$$

Now let us get a brief idea as to what these terms are:

- Bond stretching (E_{Stretch}) describes the energy of deformation of the bond length *w.r.t.* their equilibrium value. The energy near-equilibrium can be approximated by using harmonic potential, which does not allow the breaking of bonds [14]. The determination of the stretching force constant can be done using vibration spectroscopy.
- Angle bending (E_{Bend}) describes the deformation energy of the bond angles w.r.t. their equilibrium value. The energy near-equilibrium can be approximated by using harmonic potential. This force constant can be determined by vibration spectroscopic studies.

Torsional Term (E_{Torsion}) originates through space and accounts for the rotation of covalent bonds. This approximation of this term can be made with the help of a series of geometric functions.

- Fig. 1 is a representation of the non-bonded interactions that we just studied.
- Electrostatic term ($E_{\text{Electrostatic}}$) is evaluated using Coulomb's Law with the inclusion of partial charges, which are calculated by Quantum Mechanics. For better calculations, static partial charges and polarizable charges can also be considered as per one's needs.
- Van der Waals Term ($E_{\text{Van der Waals}}$)-It describes the interactive and repulsive interactions between atoms, in simpler terms, the interatomic forces. This term can be approximated by using Lennard Jones 12–6 potential, which can be thought of as a function of the distance between the centers of the two interacting atoms/molecules.
- Hydrogen Bond Term ($E_{\text{Hydrogen Bond}}$)—It describes the energy between atoms that have the potential to form hydrogen bonds. It is approximated by using 12–6 potential, which is similar to the Lennard Jones Potential, but the attractive interaction between atoms disappears faster in this case.

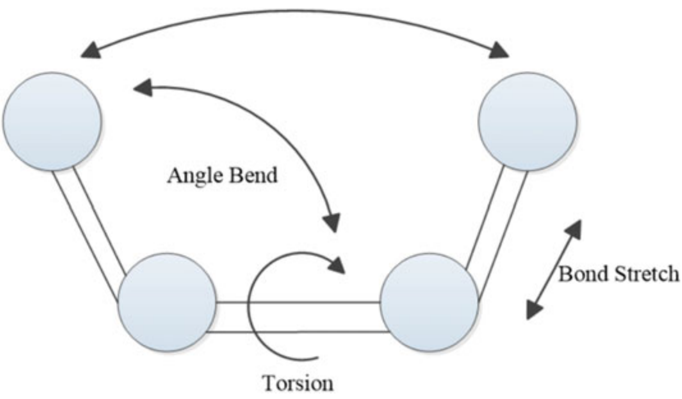


Fig. 1 Representation of the non-bonded interactions in a typical force field

- Cross terms Most interaction terms we just studied are generally not present independently in biomolecules but affect each other. Cross term accounts for all such interactions affecting others, including bend-bend, bend-torsion, stretch-bend, stretch-stretch, and stretch-torsion.

2.3 Periodic Boundary Conditions

Differential equations along with additional constraints called boundary conditions that are chosen for the approximation of a large system by using a corresponding smaller part called unit cell are known as periodic boundary conditions. Now let us understand this in simpler terms. Imagine the simulation of a system within a box-shaped container [9]. Since the system is physically fluid, it is very likely that a few particles flow out of the box due to its dynamic nature. We can apply a small trick to overcome this issue. We generate a replica of the box such that it covers the original box from all sides: whenever a particle tries to go out of the central box, there is a particle from the adjoining replica going at the same speed into the central box. The balance is maintained precisely.

Summarizing, periodic boundary conditions allow a simulation to deal with comparatively fewer particle numbers. With this operation, particles experience forces as if they are in the bulk of the liquid.

2.4 Langevin Dynamics

A real-world molecular system in a general and biomolecular system, in particular, is not likely to be present in a vacuum. Rather, it lies in such an environment, the cellular environment, where they constantly experience frictional forces. Jostling of biomolecules in such an environment causes perturbation of the system [15]. Langevin Dynamics, based on the Langevin Equation, a kind of stochastic differential equation, allows computational simulation methods to incorporate these effects. We can perceive it as an approach that necessarily imitates the solvent's viscosity but excludes its electrostatic and hydrophobic effect [9]. Suppose a system made up of N particles with mass M and coordinates X = X (t). We can use the Langevin Equation as follows:

$$MX = -\Delta U(X) - YMX + \sqrt{2\mathrm{Yk_B TMR(t)}}$$

where the notations have their standard meaning.

2.5 Time Series Calculation

To study the dynamicity of a biomolecular system based on a temporal scale, getting a proper understanding of time-dependent statistical mechanics is a vital requirement because of the recently proposed algorithms for molecular dynamics and that dynamics critically capture equilibrium time-correlation functions, particularly those corresponding to transport coefficients [16].

Components comprising time series analysis include but are not limited to the root mean square deviation (RMSD), root means square fluctuation (RMSF), surface accessibility (SA), and the radius of gyration (RGYR) [9]. These calculations help us develop an idea about the biomolecular changes that occur gradually over time. System stability during simulation is calculated by root mean square deviation. Root mean square fluctuations provide us with an overview of the residue's flexibility understudy on a determined time scale. The radius of gyration, defined as the root mean square distance of the system from its center of mass, concerns system fluctuations. Surface accessibility applies the conformation of the biomolecule so that can describe the part of the biomolecule accessible to solvent.

2.6 MD Simulation Algorithms

To study the evolution of biological systems on a temporal scale, we will use Newton's motion laws:

$$F = \frac{dP}{dT} = m\frac{d^2 r}{dt^2}$$

These classical equations of motion are integrated by applying the finite difference method. Finite difference methods are nothing but techniques applied to generate MD trajectories with continuous potential models. The basic idea behind this is that the complete integration is divided into smaller steps, so the total force of a given particle is calculated on a time scale as the vector sum of interaction between the particle being studied and other particles. Algorithms are available for integrating the equation of motion using the finite difference method, and the main assumption made by all the available algorithms is that the dynamic property obeys Taylor's theorem as follows [17–19]:

$$r(t + \partial t) = r(t) + v(t)\partial(t) + \frac{1}{2}a(t)\partial t^2 + \ldots$$

$$v(t + \partial t) = v(t) + a(t)\partial t + \frac{1}{2}b(t)\partial t^2 + \ldots$$

$$a(t + \partial t) = a(t) + b(t)\partial t + \ldots$$

where r = position, v = velocity (the first derivative of the position concerning time), and a = acceleration (the second derivative of the position concerning time).

i. Verlet algorithm

It is an algorithm commonly used for integrating the equation of motion [20]. It is a two-thirds order algorithm that applies Taylor series expansion for the position of molecule $r(t)$, one forward and the other reverse in time. Verlet algorithm employs a method that does not involve explicit velocities, and for this, it relates to an "explicit central difference method." The position of the previous step will be $r(t-d\mathrm{t})$, and to calculate a new position, we can write down the following equation:

$$r(t+\partial t)=r(t)+v(t)\partial(t)+\frac{1}{2}a(t)\partial t^2$$

$$r(t-\partial t)=r(t)-v(t)\partial(t)+\frac{1}{2}a(t)\partial t^2$$

If we add the above two equations, we obtain

$$r(t+\partial t)=2r(t)+r(t-\partial t)+a(t)\partial t^2$$

The advantages of this algorithm are that it is straightforward and self-starting, and the new positions can easily be obtained from the current and previous positions. Another advantage is that it requires less computer memory.

The disadvantage is that due to the lack of an explicit velocity term, it is difficult to obtain the velocity at the current position until the position has been computed for the next step.

ii. Velocity Verlet Algorithm

As we mentioned earlier, the Verlet method does not involve velocity. Though it is unnecessary during the actual simulation process, it is needed for the calculation of the kinetic energy in order to test the total energy conservation [21]. The advantage is that this step helps verify whether the simulation process is proceeding correctly or not. It shows a better use of the basic Verlet algorithm discussed above. The calculation of positions, velocities, and accelerations are as follows:

$$r(t+\partial t)=r(t)+v(t)\partial t+\frac{1}{2}a(t)\partial t^2$$

$$r(t+\partial t)=v(t)+\frac{1}{2}[a(t)+a(t+\partial t)]\partial t$$

Because at a particular time, all three parameters: position, velocity, and acceleration are considered, there is no compromise on the precision.

The advantage is the same as the Verlet Algorithm, i.e., it is storage efficient. The disadvantage is related to the error we can find in the range of Δt^2.

iii. Leapfrog algorithm

It was developed to control the error obtained associated with the method of Velocity Verlet. The leapfrog method calculates velocities and positions at interleaved time intervals in a way that the position r is integral time step $(t + \delta t)$ and velocity incorporates an extra half step and is defined as integral time plus a half step $(t + 1/2\delta t)$ [22, 23]. Therefore, velocities do leap over positions, and vice versa. We can write it as:

$$r(t + \partial t) = r(t) + v\left(t + \frac{1}{2}\partial t\right)\partial t$$

$$v\left(t + \frac{1}{2}\partial t\right) = v\left(t - \frac{1}{2}\partial t\right) + a(t)\partial t$$

It is a way to explicitly calculate velocities, but there is no room to simultaneously calculate positions.

As a result, it uses a different formula to estimate total energy at any point in time:$v(t) = \frac{1}{2}\left[v\left(t - \frac{1}{2}\partial t\right) + v\left(t + \frac{1}{2}\partial t\right)\right]$.

iv. Beeman's algorithm

It is very close to the method of Verlet and can be described by:

$$r(t + \partial t) = r(t) + v(t)\partial t + \frac{2}{3}a(t)\partial t^2 - \frac{1}{6}a(t - \partial t)\partial t^2$$

$$v(t + \partial t) = v(t) + v(t)\partial t + \frac{1}{3}a(t)\partial t + \frac{5}{6}a(t)\partial t - \frac{1}{6}a(t - \partial t)\partial t$$

It allows more accurate treatment of velocities and energy [24]. However, even with this method, the calculations are not perfectly made, but instead, it is a computationally expensive algorithm and therefore is not practical in the real world.

A few of the many software used for MD simulation are AMBER (Assisted Model Building with Energy Refinement), CHARMM (Chemistry at HARvard Molecular Mechanics), GROMOS (GROningen MOlecular Simulation), and NAMD (Nanoscale Molecular Dynamics).

3 Method

There are several softwares available for performing the molecular dynamic simulation of biomolecules like GROMACS, Open Babel, VMD, UCSF Chimera, etc. We can select the software of our choice and perform the task but always remember that different software uses different force fields.

MD simulations are performed in three main steps, which further consist of smaller steps: model selection; energy minimization, heating, and equilibration; and production run and analysis.

If we talk about Chimera, MD simulation can be thought of as a link to minimization and molecular dynamics routines provided by Molecular Mechanics Toolkit (MMTK), which is incorporated with it. Standard residues are assigned Amber parameters, while non-standard residues are assigned parameters using Chimera's Antechamber module.

i. Model Selection: A model system of interest should be chosen. Most of the time, complete models are available for use, but in case complete models are unavailable, the missing segments are secured, and the protonation states are conditioned. All atoms of interest should be considered and included in this step because models not included here will be ignored. The prepared molecule should be read in the pdb and psf files.

   Obtaining Files- Simulations generally begin with a crystal structure one can obtain from the Protein Data Bank. The information about atoms of use is the atom names (N, C, CA), Residue name and ID, Occupancy, Coordinates, Beta factor or Temperature Factor, and Segment ID.

ii. Energy minimization, heating, and equilibration: This step includes the equilibration of the model structure that depends on the force field of choice (T = 0). We also decide the number of equilibration steps (default 5000). Then the systemis heated by rescaling the velocities, and its stability is ensured until the system's properties stop changing with time and the system reaches a particular temperature. Preparing the system for energy minimization It includes energy searching by force-field methods, and accordingly, generating low-energy conformations. Different strategies can be used to complete the so-called minimization step, a few of which are listed below:

   - Steepest Descent—Used for highly restrained systems.
   - Conjugate Gradient—Used for large systems; applies intelligent choices of search direction; efficient.
   - Broyden–Fletcher–Goldfarb–Shanno (BFGS)—Quasi-newton variable metric method.
   - Newton–Raphson Method—Calculates both slopes of energy as well as rate of change.

Periodic Boundary conditions should be used whenever a solvent box is added. The cut-off distance should not exceed half of the smallest box dimension for maintaining periodic boundary conditions.

Fixed Atoms help specify whether one needs to freeze some atoms in a position during the calculations. Such atoms to be frozen in place are highlighted by the selection, but one must always remember that all atoms in the desired model will be considered in the energy calculations, whether they are fixed or not.

Translation Remover aids in subtracting out a global translational motion during MD and also decides which steps, by default the first, third, fifth, etc., through the end.

Rotation Remover aids in subtracting out a global rotational motion during MD and also decides which steps, by default the first, third, fifth, etc., through the end.

Topology files: It assumes that each element contains different atoms and corresponds to molecular orbital environments that result from the interaction between the atoms and their charges and orbitals. Topology files include information about atoms, atomic charges and orbitals, and atom representations in elements.

Parameter files force constants necessary to predict the bond energy, non-bonded interactions (Van der Waals and electrostatic), angle energy, and torsion energy are available in these files along with parameters proposed for energy calculations.

Solvation Some biochemical processes take place in aqueous systems and therefore, the impact of solvation is significant on the determination of molecular conformation, binding energies, and electronic properties [25]. There are two methods of model solvation: the explicit method works on solvents as being explicitly introduced to the system, while the implicit method models the solvent molecule as a continuum dielectric.

iii. Production Run and Analysis: The model is then simulated under desired conditions of NVT, NPT, etc. Finally, a production run is performed for a relevant time to get the output trajectories. The 'include production phase' helps us decide whether to include the production MD in a phase and if so, how many steps should be included. We also need to mention the time steps at which we write the trajectory files, which are further analyzed to obtain the desired properties of interest [26].

The steady advancement of potential computational sampling methods now lets us carry out the simulation process on a time scale of seconds to microseconds and even milliseconds. Here it should be significantly noted that in simulation, these millisecond scales are believed to be enormous that contradicts the in vitro experiment because, in a computer simulation, coordinates are produced at the femtosecond level. When moving from femtoseconds to milliseconds scale, we have many conformations emerged to unravel biological problems that otherwise remain unsolved.

4 Applications

MD simulations have a wide broad of applications not only in the field of biological science but in any field one can imagine ranging from physics, chemistry, biology to climatology and meteorology, video games, to film industries. Let us focus on the applications of MD simulations in biological complexes.

i. Determination of Structures and Movements of Biomolecules

   As already mentioned, we now know that the most common application of MD simulation in biomolecules is to study, analyze and mimic the flexibility, movements, and interactions of and among the different proteins. Structures determined by experimental studies by X-Ray Crystallography or NMR studies reveal only an average approximation of what the real thing could be. However, with computational simulation techniques, one could make an even more precise approximation of the types of structural fluctuations the molecules undergo. By just scrutinizing a simulation of these structures, one can quantify the movements of different regions of the molecule at equilibrium and the types of structural fluctuations that occur [27]. Such simulations also can show the dynamic behavioral properties of water molecules and salt ions, the effects of which are often critical for the proper functioning of protein and also for ligand binding.

ii. Assessment of accuracy and Refinement of modeled structures

   This method can also be used to assess the accuracy of already modeled structures or even to refine the structures built using molecular modeling techniques or experimentally in the lab. For example, it is frequently seen that experimentally determined X-Ray crystal structures are refined by a computational MD simulated annealing protocol and fit the model to the experimental data even more precisely while simultaneously maintaining a physically stable structure [28]. One advantage of this approach is that it has been shown to control model errors otherwise present. Let us consider another example. A membrane protein may suffer from artifacts due to the absence of a lipid bilayer or crystal structure suffers from such errors due to the crystal lattice packing but owing to the lucidity of the near accuracy of the simulated structures. It is now possible to correct such artifacts by performing a simulation of inappropriate solvation environments as per the requirements of the structures one is working with. Though MD simulations are extremely useful in the refinement of existing homology models, several attempts to do this have been unsuccessful [29]. MD simulations have also been applied to retrieve ensembles of conformations, against a single structure, from NMR data [30]. In each of these cases, the molecular mechanic's force field is augmented by terms that have to be taken from experimental data, which results in lower energy for structures (or structural ensembles) that are more suitable.

iii. The flexibility of Molecules
The flexibility directly modulates the association of a molecule with its neighboring atoms, molecules, and ions, and thus plays an active role in cellular function. We have already studied that the molecular dynamic system gives us clear insights into the dynamic evolution of any system. It can also be seen as reflecting its flexibility to an extent. Recently developed techniques such as Anisotropic Network Model (ANM), Elastic Network Model (ENM), Principal Component Analysis (PCA), among others, have allowed the extrapolation of prime motions in the system [31].

iv. Another interestingly important use of MD simulation is to ascertain the mechanism in which a biomolecular system will respond to perturbation. Say, for example, someone changes the molecular environment of the protein like the salt concentration or lipid composition, or adds a ligand where there was originally no ligand present or replaces a bound ligand with a different ligand, or changes the amino acid residues present in a particular protein by mutating them or by changing the protonation state of the amino acid [32, 33]. In all the cases mentioned above, simulations help one in getting a thorough understanding of the system under study. One thing to be kept in mind while performing such simulations is that one should perform it several times by using both perturbed and unperturbed systems to get clear insights into the consistent differences in the results and thus ascertain one's results.

v. Analysis of results of MD simulation of different systems helps one to answer such questions about the role of structure, flexibility, and the interactions among different biomolecules that are experimentally very difficult to address. Since simulations can occur at the scale of femtoseconds, we can observe such biological processes that occur in a jiffy, like the order in which the substructures form during protein folding [34, 35]. One can also perform a thorough study of processes like ligand binding, protein folding, conformational changes, membrane transport, etc. They also help us understand the factors controlling ligand binding and dissociation kinetics, the process of assembly of disordered proteins to form fibrils [36, 37]. Simulations may capture an entire process in one go, or they may capture it in parts, which can then be used to reconstruct the entire process [38–41].

vi. Modeling of Drug Receptor Interactions
Experimental studies help us determine the 3D conformation of ligands. Ligands, also known as drug molecules, bind to receptors, that do not have a known structure and for this, cannot be targeted directly. Instead, ligand-receptor interactions are candidates amenable to drug design. It is significant to know the structure of both the receptor and ligands before carrying out further simulations by making drug design modifications. Recent work has emphasized structure-based drug design (SBDD) and ligand-based drug design (LBDD) approaches for modern drug discovery. SBDD uses the 3D structure of drug target and free energy techniques for the task of approximating the absolute and relative binding free energy (RBFE). RBFE or alchemical approaches allow the application of MD simulations to

the initially sampled ligand and consider this first application for calculations of binding free energy differences between structurally similar ligands. This is immensely important from the point of view of drug discovery. In short, simulations are used to determine the molecule's location to bind to its receptor and how it changes the binding strength and affinity of molecules that bind elsewhere. This information, along with other geometrical, physical, chemical, and thermodynamic properties, is used to alter the structure as many times as possible to design a drug that fulfills one's needs. Once this computational task is done, the experimental scientists take over, and after its testing and approval, clinical trials take place, and if it passes the clinical trial, the drug is ready to be launched in markets.

vii. Gives Insights into Molecular Interactions on a Temporal Scale
Molecular dynamics simulation generates pictures of atomic-level details of the dynamic evolution of the biomolecular system. This property clubbed with temporal scale for MD simulations, enables us to predict different feasible cellular interactions and behaviors based on which modifications in existing structures can be made and seen and is immensely helpful in studying the properties of these samples, which can be used as potent drugs in the market.

viii. Docking
Docking is the process by which two or more molecular structures orient themselves so that they bind to each other to form a stable complex. MD simulation techniques offer approaches for monitoring different types of interactions, including DNA-ligand, DNA–protein, protein–ligand, and protein–protein. More interestingly, they enable us to probe emerging types of molecular interactions that participate in the formation of more complex structures. For this reason, docking is performed first, and then MD simulation is done so that one can know the effect of interactions on a temporal scale. Docking software like Glide, GOLD, AutoDock, etc., make use of different algorithms to calculate a docking score based on different parameters like the surface of contact, electrostatics, etc. [42], and a good score is considered as one which has a good binding affinity. Then, experimental and clinical studies are necessary to test the findings in a realistic platform, although it is a challenging task in terms of time and cost to obtain all the docked structures. This calls a need to filter out structures based on thermodynamic and structural characteristics as explained by MD simulation techniques.

ix. Protein folding
It is an interesting topic in biology. Though the 3D structure of proteins has been studied pretty well, less is known about protein folding. MS simulation could help to identify the folding mechanism using computational power. For example, the MD simulation of a sub-domain of villin protein was done, which gave significant insights and a glimpse of hope toward the proper understanding of the protein folding mechanism [43].

x. To understand how mutation affects Interactions
To understand and realize the effects of a given amino acid and its binding, a mutation is introduced to the residue and the difference in the simulated trajectories both before and after the process is studied well. The ligand can also undergo refinements and modifications so that its affinity is improved and insights into its structural properties and interactions become certified [9].

5 Future Scope

As students of science, we all know that the macroscopic properties of elements owe a great deal to the time-dependent underlying microscopic properties and interactions among atoms and molecules. Molecular dynamics simulation has unbolted an incredibly huge number of doors for the research enthusiasts in biological sciences, from the study of protein folding to the physics and chemistry behind the interactions among biomolecules, molecular docking to drug design, etc. The entire technique of MD simulation relies solely on the trustworthiness of the model, force field calculations, and the thermodynamic property calculation, and the ability of particular software to be able to mimic a process with as much reality as possible. Even though so many computer simulation techniques have been developed, there is always the scope of improvement, and simulation techniques do not show cent percent accuracy in their results, so better and more accurate techniques can always be developed. With better results of the simulation process, the foundation for future studies on ion exchange is being laid. Also, we need to develop more robust algorithms and one with a shorter number of steps. There is also scope for the development of lightweight and free software. Also, algorithms that are computationally less intensive are the need of the hour.

6 Conclusion

MD simulation technique is now more than sixty years old, yet it is such an incredibly amazing technology that there is still much excitement in the scientific community about this technology, and also it has maintained its spot in the limelight since then. But it was only recently that MD achieved time scales compatible with that of the biological system. Today, conformational changes and assembly of proteins and ligand–protein or protein–protein or ligand-ligand interactions can be studied with such ease with effective simulation. It also provides knowledge about the interactions at the atomic level, which directly affect the functions and behavior of the molecule, uncovers how different energy components make different contributions to molecular binding and stability. Therefore, it can rightfully be said that molecular dynamic simulation is a technique which is the best instance of productive research being carried out where scientists from the background of Physics, Chemistry, Biology, and Computer Science come forward and work together on a common platform and is thus an exemplary technique developed in the era of interdisciplinary research.

Core Messages

- Molecular dynamic (MD) simulation offers computer-aided techniques to apprehend the dynamic behavior of biomolecules by visualizing atoms and molecules when interacting with each other over a period of time and analyzing their physical movement and chemical interactions.
- MD simulation aims to predict the behavior and movement of atoms as a time-dependent function with the hope of comprehending the properties of molecules concerning their structure and conduct under different conditions.
- Though Alder and Wainwright first introduced MD simulation in 1957–1959, the first simulation dates back to 1964 when Rahman et al. initiated developing real-world liquid argon.

Acknowledgment We would like to acknowledge Director (Admin), **Mr. Syed Sharique Hussain**, Salfia Paramedical Institute Darbhanga, (SPI) Darbhanga, (director_admin@salfiainstitute.edu.in), **Dr. Naima Saman** (pursuing MD), **Dr. Syed Abdul Hakeem**, **Mrs. Siddiqua Khatoon**, all members of Farid/Salfia family, Faridia Hospital, **Dr. Faiyaz Ahmad**, Chairman, Madhubani Medical College, Madhubani, **Dr. Faraz Fatmi** (JD(U)), **Mr. Ashraf Ali Fatmi** (JD (U)), **Dr. Shakeel Ahmad** (Congress), **Mr. Salman Akhtar**, Chief Functionary, Nalanda Minority Educational and Welfare Trust, **Dr. Muhammad Nezamuddin (Jawed)** and **Dr. Arifa Nishat**, Ruston, LA, USA, **Mr. Prashant Kishore**, (I-PAC), **Dr. David K Mills** and **Dr. Thomas C Bishop** (Louisiana Tech University, Ruston, LA, USA), **Dr. Manzoor Alam**, Chairman, AIMC, Brother Aman, Aligarh Bachchon Ka Ghar, Aligarh, **Hafiz Hassan Badar**, Secretary, Hazrat Umar Farooque Academy, Nazra-Ranipur, **Mrs. Shahnaz Badar**, and motivator for Abdur Razzaque Family, **Mr. Saeeduzzafar**, (IDB Jeddah) and my life partner, **Mrs. Rana Kamal Sufian**.

References

1. Durrant JD, McCammon JA (2011) Molecular dynamics simulations and drug discovery. BMC Biol 9(1):1–9
2. Ma B, Kumar S, Tsai C-J, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12(9):713–720
3. Ma B, Shatsky M, Wolfson HJ, Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 11(2):184–197
4. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. Ber Dtsch Chem Ges 27(3):2985–2993
5. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discovery 2(7):527–541

6. Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Sci 8(6):1181–1190
7. Feynman RP, Leighton RB, Sands M (2011) The Feynman lectures on physics, Vol. I: the new millennium edition: mainly mechanics, radiation, and heat, vol 1. Basic books
8. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99 (6):1129–1143
9. Karaa WBA (2015) Biomedical image analysis and mining techniques for improved health outcomes. IGI Global
10. González M (2011) Force fields and molecular dynamics simulations. École thématique de la Société Française de la Neutronique 12:169–200
11. Jorgensen WL, Tirado-Rives J (1996) Monte Carlo versus molecular dynamics for conformational sampling. J Phys Chem 100(34):14508–14513
12. Ishida H, Agag T (2011) Handbook of benzoxazine resins. Elsevier
13. Mayo SL, Olafson BD, Goddard WA (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem 94(26):8897–8909
14. Pearlman DA, Kollman PA (1991) The overlooked bond-stretching contribution in free energy perturbation calculations. J Chem Phys 94(6):4532–4545
15. Tamar S (2002) Molecular modeling and simulation. Interdisciplinary applied mathematics. Springer, NY
16. Gullingsrud JR, Braun R, Schulten K (1999) Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. J Comput Phys 151(1):190–211
17. Akimov AV, Kolomeisky AB (2011) Recursive taylor series expansion method for rigid-body molecular dynamics. J Chem Theory Comput 7(10):3062–3071
18. Kutteh R (1999) New methods for incorporating nonholonomic constraints into molecular dynamics simulations. J Chem Phys 111(4):1394–1406
19. Streett W, Tildesley D, Saville G (1978) Multiple time-step methods in molecular dynamics. Mol Phys 35(3):639–648
20. Grubmüller H, Heller H, Windemuth A, Schulten K (1991) Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul 6(1–3):121–142
21. Martys NS, Mountain RD (1999) Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. Phys Rev E 59(3):3733
22. Toxvaerd S (1991) Algorithms for canonical molecular dynamics simulations. Mol Phys 72 (1):159–168
23. Van Gunsteren WF, Berendsen HJ (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1(3):173–185
24. Van Gunsteren W, Berendsen H (1982) Algorithms for Brownian dynamics. Mol Phys 45 (3):637–647
25. Rosenthal SJ, Jimenez R, Fleming GR, Kumar P, Maroncelli M (1994) Solvation dynamics in methanol: experimental and molecular dynamics simulation studies. J Mol Liq 60(1–3):25–56
26. Jabbarzadeh Kaboli P, Ismail P, Ling K-H (2018) Molecular modeling, dynamics simulations, and binding efficiency of berberine derivatives: a new group of RAF inhibitors for cancer treatment. PLOS ONE 13(3): e0193941
27. Miao J, Hodgson KO, Sayre D (2001) An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. Proc Natl Acad Sci 98(12):6641–6645
28. Terwilliger TC, Read RJ, Adams PD, Brunger AT, Afonine PV, Grosse-Kunstleve RW, Hung L-W (2012) Improved crystallographic models through iterated local density-guided model deformation and reciprocal-space refinement. Acta Crystallogr D Biol Crystallogr 68(7):861–870
29. Mirjalili V, Feig M (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. J Chem Theory Comput 9(2):1294–1303

30. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433(7022):128–132
31. Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci 106 (34):14349–14354
32. Fribourg M, Moreno JL, Holloway T, Provasi D, Baki L, Mahajan R, Park G, Adney SK, Hatcher C, Eltit JM (2011) Decoding the signaling of a GPCR heteromeric complex reveals a unifying mechanism of action of antipsychotic drugs. Cell 147(5):1011–1023
33. Peng Y, McCorvy JD, Harpsøe K, Lansu K, Yuan S, Popov P, Qu L, Pu M, Che T, Nikolajsen LF (2018) 5-HT2C receptor structures reveal the structural basis of GPCR polypharmacology. Cell 172(4): 719–730. e714
34. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334(6055):517–520
35. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420(6911):102–106
36. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127(2):476–477
37. Nguyen HD, Hall CK (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proc Natl Acad Sci 101(46):16180–16185
38. Bernardi RC, Melo MC, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochimica et Biophysica Acta (BBA)-General Subjects 1850(5): 872–877
39. Harpole TJ, Delemotte L (2018) Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. Biochimica Et Biophysica Acta (BBA)-Biomembranes 1860(4):909–926
40. Hertig S, Latorraca NR, Dror RO (2016) Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. PLoS Comput Biol 12(6):e1004746
41. Schwantes CR, McGibbon RT, Pande VS (2014) Perspective: markov models for long-timescale biomolecular dynamics. J Chem Phys 141(9): 09B201_201
42. Kroemer RT (2007) Structure-based drug design: docking and scoring. Curr Protein Pept Sci 8(4):312–328
43. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282(5389):740–744

Mohammad Sufian Badar PhD has served as a Teaching Faculty in the Department of Bioengineering at the University of California, Riverside, CA, USA. He served as an Analytics Architect in CenturyLink for more than a year in Denver, CO, USA. Currently, he is an Director (Academic), Salfia Paramedical Institute Darbhanga, India, (www.salfiainstitute.edu.in). He possesses an excellent academic record with an MS degree in Molecular Science and Nanotechnology and Ph.D. in Engineering from Louisiana Tech University Ruston, LA, USA, respectively. Before joining the Ph.D. program at Louisiana Tech University, he graduated with an MSc in Bioinformatics from Jamia Millia Islamia University, New Delhi, India. Dr. Sufian has over 14 years of teaching, research, and industry experience. He has published his research in conferences and high reputed International journals. He has authored many chapters in the area of Artificial Intelligence/Machine Learning, Blockchain/IoT. Currently, he is Editing two books: 1. Diagnosis & Analysis of

COVID-19 using Artificial Intelligence and Machine Learning based Techniques, 2. COVID- 19: Cause, Transmission, Diagnosis, and Treatment. He has developed an algorithm for Face Detection, Recognition, and Emotion Recognition. He is currently in the process of developing a device that, using Biosensors, can correlate the physiology of the human body with the emotion recognition algorithm, giving us a clear measure of the amount of stress hormones in the body. Currently, my group and I have developed an ML model which predicts Covid infection based on symptoms only, and we are now working on increasing the accuracy of our model.

Shazmeen Shamsi has an M.Sc. degree in Bioinformatics from Jamia Millia Islamia, New Delhi, India, 2020 and B.Sc. (Honors) in Botany from the University of Calcutta, West Bengal, India, in 2018. She has a Diploma in PC Application, and certifications in "Blockchain & Applications" and "Python Programming". She has worked as a research intern in the School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi, and has co-authored three book chapters. She is presently a PGD (Big Data Analytics) student at Jawaharlal Nehru University, New Delhi, and is also working as a Research Trainee under Dr. Mohammad Sufian Badar, Senior Teaching Faculty, Department of Bioengineering, University of California, Riverside, CA, USA. Her current research interests include Artificial Intelligence, Internet of Things, Network Security, Data Analytics, Social Networking, and Simulation Techniques used in Bioinformatics.

The World Big Challenges Faced by Nanoscience: Examples of How Integrated Science Can Save the World

8

Marco Carofiglio, Marco Laurenti, and Valentina Cauda

> "*Arranged one way, atoms make up soil, air, and water arranged another, they make up ripe strawberries. Arranged one way, they make up homes and fresh air, arranged another, they make up ash and smoke.*"
>
> Kim Eric Drexler

Summary

Our world has plenty of challenges, concerns, as well as human and environmental problems. One cannot solve most of them, but constructive human cooperation is essential. Science and engineering can also contribute, applying a broad and multidisciplinary approach. Nanotechnology deals with the creation and manipulation of matter on a scale ranging from 1 to 100 nm. This science field that emerged in the late'50 s has revolutionized the comprehension of matter and opened the door to a plethora of novel, unprecedented, and intriguing applications and solutions to current problems. This chapter aims to overview the most urgent world challenges of our era, potentially solved by employing nanotechnologies. The idea is to trace a direction to which all from

M. Carofiglio · M. Laurenti · V. Cauda (✉)
Department of Applied Science and Technology, Politecnico Di Torino Corso, Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: valentina.cauda@polito.it

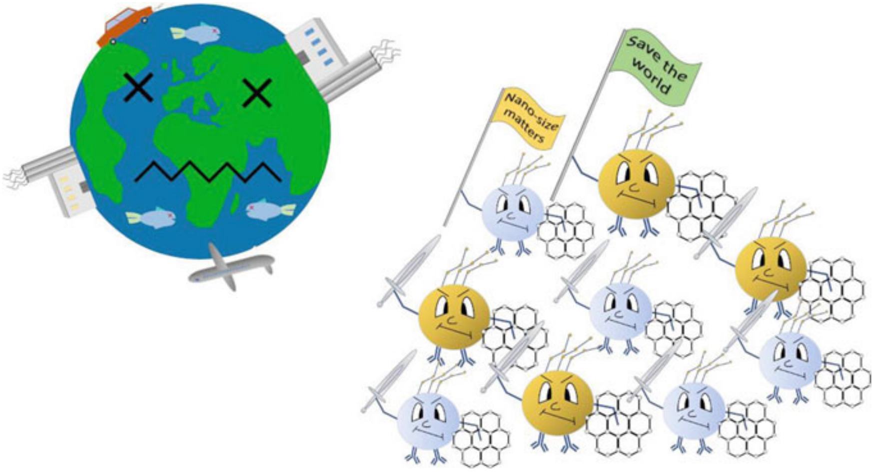
M. Carofiglio
e-mail: marco.carofiglio@polito.it

M. Laurenti
e-mail: marco.laurenti@polito.it

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Turin, Italy

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_8

scientists, professionals, and experts, to industrial people, policymakers, young people, and students can recognize with their professionality and create a common conscience and integrity on where we are and where we are going.



Some of the most urgent problems of our modern world can be solved by employing nanotechnologies and all related nanomaterials involving sciences.

The code of this chapter is 01001110 01100001 01101110 01101111 01110011 01100011 01101001 01100101 01101110 01100011 01100101.

1 Introduction

Can Nanotechnology solve the biggest world problems? We open this chapter with this challenging question and seek to answer which will be not univocal. For this, we have first analyzed the most significant challenges of our contemporary world and which current solutions have been yet available to face them. Then we have pushed our look on how the nanoscience-related disciplines can help, contribute or completely solve these problems.

As far as climate changes, clean water supply, poverty elimination, health for everyone, availability of life-saving pharmaceuticals, international security, and fight against terrorism, renewable energies are concerned. All these challenges and many others are dramatically actual and present in our everyday life. Most of them can only be solved through human cooperation, science, and engineering, typically with a multidisciplinary approach, touching every scientific discipline from biology to quantum physics. Such a multidisciplinary approach is the basis of nanotechnology, i.e., the science of ultrasmall objects (from 1 to 100 nm in size). The creation and manipulation of materials at this dimensional scale have revolutionized the matter's comprehension, opening the door to a plethora of novel, intriguing

Table 1 The World's greatest challenges in author's vision

CHALLENGE N.	WORLD CHALLENGE
1	**Renewable Energies: Cheap Energy**
	Solar energy: new materials for DSSC or alternatives
	Efficient energy conversion, recovery, and storage
	Wearable electronics for Health monitoring
2	**Remove pollution, clean natural resources**
	Clean and Drinkable water
	Carbon-based nanomaterials
	Nanoparticles for water disinfection
	Semiconductor photocatalyst
	Facing Air and Soil Pollution
	Nanomaterials for CO_2 capture and reduction
	Air/soil purification from other pollutants
3	**Health for everyone**
	Tissue regeneration
	New solutions for replacing antibiotics
	Fight cancer disease – new diagnostics for early detection of cancer or circulating tumor cells
4	**Restoration and Conservation of Public Heritage**
	Protective films and materials for conservative restoration
	Protection and restoration of paintings
5	**Reduced Impact in Human Mobility**
	Improved aerospace vehicles
	Aerogel for aerospace applications
	Carbon nanotubes in aeronautics

properties, from quantum physical effects to broad surface area availability. Many atoms in nanosized materials are at the surface, leading to increased surface chemical reactivity and surface-area-to-volume ratio. Nanomaterials thus show an amplified capability to interact with other species and the environment. Therefore, unique, unprecedented, and novel applications could arise from these new properties not present in the bulk material. Considering the possible technological solutions, we have identified five world's big challenges, as highlighted in Table 1. For clarity and completeness, each one is articulated in sub-challenges. Below are the possible nanotechnological solutions currently available for some of them.

We do not forget the main world's problem of eliminating discrepancies among humans: religion, minorities and racial injustices, wars, rich/poor, and societal differences. However, nanotechnology cannot face such huge societal problems, except trying to solve parts related to clean accessible water and health for everyone, separately discussed as single challenges. As it is clear from the following discussion, cutting-edge engineering techniques are central to promote innovation in various fields of science and technology. Personal skills like passion, determination, engagement, and an enthusiastic attitude are determinants in this field. Still, they are innate or have to be instilled in young scientists and engineers to produce innovative results facing global challenges. It is of note that, while the World Biggest Challenges require engineers and scientists to solve them, one of the most prominent challenges that remain is increasing the number of young engineers and build a multifaceted educational approach. As a further issue, we wish to recall that such scientific fields are still male-dominated. That is, while female engineers are an under-represented fraction, new high-school and further high-education models are still needed to empower the female portion among scientists and engineers.

2 Challenge N. 1: Renewable Energies: Cheap Energy

Energy occupies as always, a prominent position in the economy and development, both nationally and worldwide. In an energy context in which fossil fuels, increasingly in demand, run out quickly while greenhouse gas emissions are increasing, renewable energies and the development of related technologies show a fundamental role.

2.1 Solar Energy: New Materials for DSSC or Alternatives

Photovoltaic systems that use solar energy have gained control over about 2% [1] of the world's electrical power. As stated by the International Energy Agency (IEA) predictions, solar radiation's contribution will rise to 9% in 2040. It will undoubtedly lead to strong growth in industrial and research fields related to this area. The main challenge still relies on improving photovoltaic solar cell efficiencies and, more in general, in implementing innovative technologies adaptable to power buildings, transportation, and commonly used devices, i.e., small portable and wearable electronics.

At present, one can distinguish among three different solar cell generations, which account for the technological development achieved so far. Mono and polycrystalline silicon-based photovoltaic cells belong to the first generation with an efficiency that varies between 19.9 and 26.7% [2]. Today about 80% [3] of the operating photovoltaic panels belong to the first generation (Fig. 1).

The second generation occurred in the'70 s and works on thin-film technology. Materials commonly used are cadmium telluride (CdTe), indium gallium copper selenide (CIGS), and amorphous silicon (a-Si), deposited on substrates (glass, polymers, or metals). This technology allowed to reduce the costs of raw materials and improve their mechanical properties.

The third (and most recent) generation is related to the challenge of obtaining high-efficiency and low-cost solar cells. This generation exploits many advantages deriving from nanotechnology and includes dye-sensitized solar cells (DSSC),

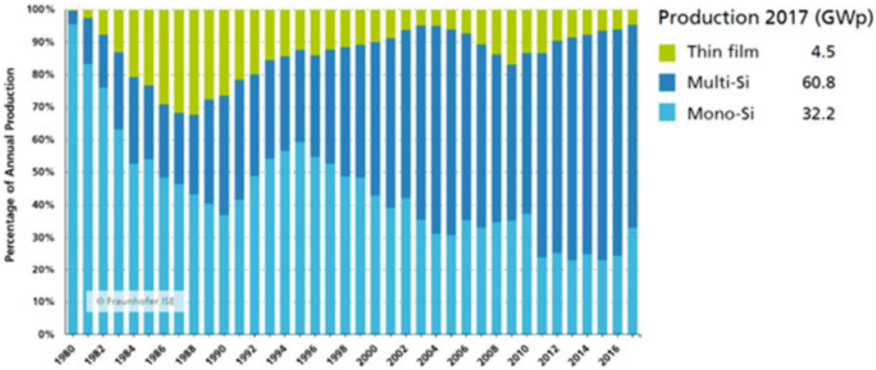


Fig. 1 The market of photovoltaic technologies, actually dominated by mono- and polycrystalline silicon photovoltaic solar cells [4]

multi-junctions, and quantum dots/wires, still in the research/prototype/testing phase. The challenges to be faced are manifold, particularly linked to the efficiency (which continues to be relatively low) and the choice/processing of raw materials in a low-cost perspective. In 1991, O'Regan and Grätzel [5] built the first nanocrystalline solar cell, whose photoelectric conversion efficiency was 7.1%. The results of this research have allowed DSSC to represent a valid alternative compared to other technologies. DSSCs have a series of advantages, including low production costs, flexibility, lightweight, energy recovery in less than a year (the so-called energy payback), two-faced cells capable of capturing light from all angles. In addition to direct light, DSSCs can absorb diffused light, thus significantly increasing the energy production hours during the day. Consequently, the use of DSSCs has not been limited to Architecturally Integrated Photovoltaic (FAI) applications. Still, they have applications in indoor (partition walls, doors, furnishing elements, etc.) and electronic devices.

The Grätzel cell consists of two transparent electrically conductive surfaces, cathode and anode, typically made by glass slides coated with fluorine-doped tin oxide (FTO) conductive thin film. A layer deposited on the anode typically consists of a nanostructured or nanoporous crystalline semiconductor, like titania (TiO_2), zinc oxide (ZnO), or tin oxide. The aim is to offer a broad surface area for dye adsorption, which can absorb solar light frequencies and inject the excited electrons into the semiconductor's conduction band. Additionally, the semiconductor's crystallinity and particle morphology should promote the photogenerated electrons' separation from holes and inject them into the anode toward the external circuit. A conductive layer covers the cathode as thin films of carbon or metal layers. In this regard, the use of graphene coated with metal nanoclusters or nanosheets (like platinum, Pt) can benefit from the high catalytic surface area of Pt coupled to the decrease in electronic transfer resistance due to graphene's high conductivity [6].

The selection of a suitable dye is also an important parameter. In particular, the challenge is to increase the absorption range of sunlight and the efficiency of converting light into electricity. In traditional DSSC, the dyes typically used are Ruthenium-based, like N3 and black dyes. These are organic molecules coordinated by the Ru metal center and present carboxylic groups to bind to the semiconductor layer through chemical adsorption. Nanotechnology shows that a valid alternative to organic dyes in DSSC can be Quantum Dots (QDs) [7]. QDs are photosensitive semiconductor nanoparticles with very low size ranges, from 2 to 7 nm. They show discrete energy levels of their conduction band to which photogenerated electrons can be excited. QDs can be deposited on the semiconductor-based anode as they form colloidal solutions in various organic solvents. QDs can absorb photons at specific wavelengths (from UV to visible range) according to their diameters. The production of QDs of various sizes thus allows having access to the broad range of solar radiation having the maximum conversion into electricity.

2.2 Efficient Energy Conversion, Recovery, and Storage

Nowadays, smart, small, and even wearable electronic devices are emerging to achieve the promise of green energy production, conversion, and storage. Mechanical movements of vehicles, vibrations, human motions, and even physiological involuntary movements (like heartbeat or respiration) can be sources of mechanical energies that can be converted into electrical energy and stored. The field of energy nanogenerators has recently emerged to provide devices able to convert mechanical energy (from compressive or bending deformations, vibrations, human or automotive motion) into electric power. There are several examples of energy nanogenerators relying on different working principles. The triboelectric nanogenerator (TENG) is an innovative technology recently specialized in energy-harvesting. It can convert mechanical energy into electric energy. It works on the coupling between contact-electrification and electrostatic induction. TENG can both scavenge small vibrations of mechanical energy and generate large-scale energy as well. Thus, it is an efficient technology for various energy scale applications [8], associated with low-cost. It might be taken as environmentally friendly and universally available, making the design of highly flexible and wearable TENG feasible. Such means can act as energy harvesters from human motions and as sources of power for wearable electronic devices. Today triboelectricity is ubiquitous, opening up the enormous possibility to select materials for TENG device fabrication. For wearable energy harvester applications, textile-based TENGs can be produced. For this particular purpose, such materials seem to meet various criteria. Lightness, softness, washability, breathability, and stretchability for being wearable like a daily cloth are key, among others [9].

Another charging principle of energy nanogenerators is piezoelectricity. Piezoelectric potential-driven nanomaterials include PZT, ZnO [10, 11], barium titanate [12], and piezoelectric polymers like polyvinylidene fluoride [13, 14]. The energy generated by a flexible piezoelectric composite nanogenerator is highly applicable. It can particularly recover mechanical energy by storing it in a next-generation supercapacitor for further use [15].

There are smart batteries developed to store electrical energy generated from various sources efficiently. They are available for various purposes, like automotive, electronic devices (PC, smartphones, smartwatches), wearable electronics. A Smart Battery (SmB) consists of a battery device equipped with onboard electronics that can enable various operations, called Smart Battery Data. It allows estimating battery parameters, conduct battery performance analysis, as well as alarm conditions detection, modulation of battery charging algorithms, and communication with other SmBs devices. The main issues which should be faced when designing and fabricating an SmB are battery capacity and power relative to volume, weight, mechanical properties, heating, and costs; battery life; toxicity of constituting elements and compounds; recharge time and type of technology.

Due to the large spread of devices that use SmB, it is important to fix all the issues mentioned above to improve their performances. Most engineering and research efforts are devoted to increasing the power capacity and reducing the charging time.

As an industrial solution, lithium-ion batteries (LIBs) are common, and new usages are still emerging. The largest market consists of portable electronic devices, followed by road transportation and aerospace applications, as seen in satellites and aviation. The LIBs technology has various applications, for example, in power supply systems, off-grid and grid-connected, and medical devices, where many improvements in the last years have occurred, such as in hearing aids, biometric monitoring systems, and so on.

LIBs can overcome many of the quoted problems mentioned above. LIBs show some disadvantages, for example, the high initial cost, the requirement of an advanced battery management system, non-optimal flexibility, and the need for frequent charging. Furthermore, there are also safety concerns about LIBs, possible thermal runaway incidents, weak recovery, and difficult or complex recycling schemes. The scientific research currently focuses on the use of nanomaterials to improve LIBs. An interesting field is creating flexible batteries using cellulose or polyolefin membranes applicable to commercial separators in SmB [16]. The aim is to produce flexible batteries in miniaturized or wearable electronics to achieve high comfort and safety. Nanomaterials can also be employed as functional elements and/or supports to improve batteries' capacity and charge time. They can be used as battery electrodes and current collectors. Carbon nanotubes (CNT), lithium nanofibers, and two-dimensional materials like graphene or MXene (2D transition metal carbides and nitrides) have great applicability due to their high surface area, superior electrical conductivity, low lithium-ion diffusion energy barrier, and mechanical properties of interest [17]. 3D MXene foams also bring better electrochemical performance to the battery. It is a highly porous and flexible material that provides massive active sites, improves lithium storage capacity, and facilitates Li-ions transfer. Performances of 3D MXene foams are far better than 2D layers. An MXene foam can even be used as an electrolyte, as it provides massive channels for ion transfer [17].

There are electrolyte nanomaterials proposed in the form of nanostructured gel polymers. Liquids cannot be generally used because they cannot prevent the growth of lithium dendrites. To overcome this issue, a gel polymer electrolyte can be used.

For example, using a SiO_2-GPE membrane, a synergistic effect is obtained, i.e., a fully compatible interface between the electrode and the electrolyte interface and Li's growth inhibition dendrites, thus showing superior electrochemical performance [18]. Furthermore, new technology is emerging for producing a solid electrolyte-polymeric matrix (SEPM) as an electrolytic layer. A crosslinked and highly porous polymer matrix can fill the pores with an organic polymer (polyimide), making them mechanically robust while providing high transport rates of Li^+ cations among the solid electrolyte. Despite increasing the composite's mechanical hardness, this networked structure further provides self-healing properties while producing small effects on the electrolyte ionic conductivity [19].

Nanomaterials can add lifetime and charge to a battery but not without compromises; flexibility is core for a smart battery in the wearable electronics world, but even if there are plenty of solutions to enhance this property, it is also true that every possible improvement leads to a loss in capacity and energy density.

This problem can be avoided by focusing not only on the battery performance but also on introducing an external component like a self-charging power unit. By combining an energy nanogenerator (as those mentioned above) with a LIB, a self-sufficient power system can be obtained. For example, it is possible to integrate a whole-textile TENG-cloth and a flexible LIB belt. For this purpose, conventional polyester textiles, which are flexible and insulating, were converted into a conductive material after electroless plating with a conformal Ni film. The modified material was then used as electrodes in the TENG-cloth and as current collectors in the LIB [9].

For piezoelectric nanogenerator-based self-charging LIBs, a polyvinylidene fluoride (PVDF)–PZT nanocomposite piezo-separator is used to create the piezoelectric field along the z-axis when the battery is compressed by a uniaxial force (Fig. 2). The electrical energy is then used to charge the Li-Ion battery creating a self-charging power cell (SCPC).

With a compressive force of 10 N at a frequency of 1.5 Hz applied to the SCPC, one can expect an increase in the device's voltage from 210 to 297.6 mV in the 240 s. After that, the device takes 37 s to discharge back to its original voltage with 1 μA of constant discharge current [21].

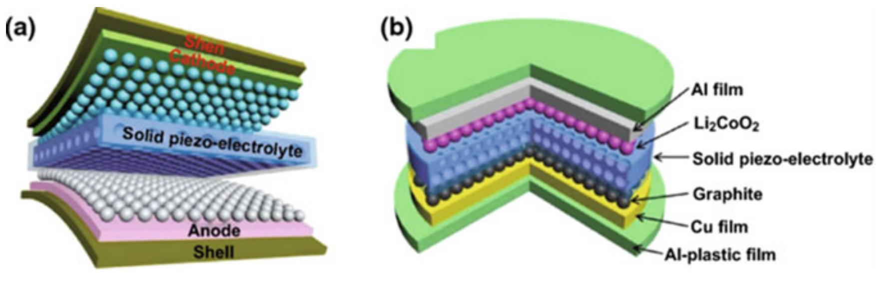


Fig. 2 **a** All-solid-state self-charging power cell (SCPC); and **b** a flexible SCPC where the mechanical compressive energy can be used to generate electric charge thanks to a nanogenerator and then stored in a Li-ion battery. (Adapted with permission from [20])

2.3 Wearable Electronics for Health Monitoring

Microcircuits and nanosized components can help with the design of portable monitoring electronic devices. They can exploit the quantum tunneling effect of electrons to transport current or the emission of light stimuli. These effects require low energy to be activated. Therefore an advantage over macroscopic solutions is lower energy expenditure. Also, the high surface area provides greater sensitivity to pressure and allows obtaining high-resolution data or images. Such contributed to the development of devices, such as ultra-flexible and pressure-sensitive pedobarographic insole [22, 23] and an electroluminescent skin (or e-skin) device [24]. The e-skin or bionic skin consists of ultra-thin (3 μm-thick) and ultra-flexible polymeric layers containing integrated circuits, to be applied to the human skin. There are several possibilities of health monitoring, depending on the nanocomponents and used materials:

- Ultra-flexible organic photonic skin: this very thin electronic film applied to the skin allows monitoring the patient's health. It is an oximeter in which polymeric diodes emit light of three different colors and organic photodetectors. Thanks to their presence, the device not only acts as a sensor but is also able to communicate data via a display on the surface that changes color based on the type of detected data [25]; and
- Ultraflexible and pressure-sensitive electroluminescent skin: includes cellulose and nanowire nanohybrid network (CNN), i.e., a composite material with high piezoresistive properties, and QLED that allow the emission of light thanks to the presence of quantum dots. This device enables viewing pressure images in real-time with a very high resolution [26]. Through this device, many parameters can be measured avoiding invasive procedures: from the detection of vital functions to the monitoring of tumor nodules, passing through the increase in tactile sensitivity. With this solution, hospital monitoring can be avoided, and constant control of the parameters can be obtained.

3 Challenge N. 2: Remove Pollution, Clean Natural Resources

The research for environmental pollution solutions is among the most compelling challenges that nanoscience is dealing with nowadays. Among the different approaches, novel nanomaterials' design for removing pollutants (carbon monoxide, heavy metals, hydrocarbons, volatile organic compounds, etc.) from water, air, and soils are noteworthy. In this regard, nanomaterials and nanotechnologies can significantly contribute to creating advanced materials and technologies for the remediation and purification of water, air, and soils.

3.1 Clean and Drinkable Water

The dramatic reduction in natural water resources is strongly motivating the exploitation of nanomaterials and nanotechnologies for producing clean and drinkable water. This goal can be achieved by removing toxic compounds from seawater (water desalination), lakes, rivers, and groundwater. Nowadays, water purification technologies are often based on polymeric membranes to capture undesirable compounds (salts, drugs, pollutants from industries, etc.). Despite the good efficiencies and low production costs, polymeric membranes suffer from some major drawbacks such as low salt rejection, poor water flux, and reduced chemical stability in aqueous environments, which causes fouling phenomena. All these aspects require replacing the membrane frequently and increase the energy demands and cost associated with the water treatment. On the other hand, nanomaterials with enhanced affinity and selectivity toward specific contaminants demonstrated great potential in membrane technologies.

3.2 Carbon-Based Nanomaterials

Water desalination using carbon nanotubes (CNTs) has been widely investigated, either employing CNTs as self-standing membranes or reinforcing fillers in polymer-based composites. In both cases, important enhancements in water permeability and salt rejection were demonstrated [27]. Thanks to the one-atomic-layer thickness, excellent mechanical strength, and chemical stability, nanoporous single-layer graphene recently found successful application as a new generation membrane able to separate organic (salts, drugs, oils) and inorganic (heavy metals) compounds from water. This can be obtained if controlled nanopores (diameter <1 nm) are opened within the graphene structure. Various methods have been reported for this scope, such as the exposition to O_2 plasma reactive atmosphere [28], oxidative chemical etching, and focused electron/ion beam [29, 30]. Depending on the method, the pores' diameter ranges from the subnanometer (plasma treatments) to the nanometer (ion/electron beam and oxidative chemical etching) scale. In the former case, nanoporous graphene-based membranes for water desalination with 90% salt rejection have been successfully fabricated [31]. While membranes suitable for separating bigger and toxic compounds (dyes and oils) from water were demonstrated in the latter cases [30]. Due to the one-atomic thickness, single-layer graphene cannot be considered a self-supported membrane at all, and it needs macro/microporous membrane support. Despite the promising results, large-scale production of nanoporous graphene membranes is still far due to the difficulties in synthesizing single-layer graphene over large areas and implementing techniques that allow obtaining a controlled and reproducible induced porosity. Graphene oxide (GO) can be obtained by the Hummers method, i.e., the oxidation of graphite crystals with a mixture of sulfuric acid, sodium nitrate, and potassium permanganate, and it allows the fabrication of new membranes with superior properties. GO membranes have a multilayer structure made of GO flakes stacked one to each other with a subnanometer interlayer spacing and kept together by a hydrogen bond. This peculiar structure makes GO membranes very promising for molecular sieving, and their separation performance can be customized by tuning the GO interlayer spacing [32]. GO has also been used as a nanofiller and the composite GO/polymer membranes showed important improvements in terms of water flux and surface hydrophilicity [33].

3.3 Nanoparticles for Water Disinfection

Several nanomaterials show good antimicrobial properties thanks to reactive oxygen species production that deteriorate bacteria components, inhibit enzyme activities and DNA synthesis, or interrupt energy transduction. Among nanomaterials, titania (TiO_2), ZnO, and Ag nanoparticles (NPs) have been proposed as the best candidates because of various advantages, including good water stability, low costs, and non-toxic behavior. For example, Ag NPs with an average diameter lower than 10 nm showed a highly toxic behavior against *Escherichia coli* and *Pseudomonas aeruginosa* [34]. Triangular Ag NPs exhibited a superior antimicrobial activity than Ag nanorods and Ag nanospheres, emphasizing a key role of particle size and shape in maximizing the desired antibacterial effects [35]. Similar findings have been observed for ZnO [36] and TiO_2 nanomaterials [37].

3.4 Semiconductor Photocatalyst

Thanks to their optical and electrical properties, semiconductors can convert absorbed light into free electrons (photoelectrons), which react with the surrounding aqueous environment and generate reactive oxygen species (ROS). ROS may oxidize organic pollutants and induce their degradation into non-toxic products. The process described above is called photocatalysis. TiO_2, ZnO, and tungsten oxide (WO_3) are the most investigated semiconductors for this purpose [38]. The semiconductor mentioned above photocatalysts allowed removing various pollutants from wastewater, including organic dyes, benzene, toluene, ethylbenzene, and xylene (BTEX).

4 Facing Air and Soil Pollution

The capture and storage of carbon dioxide (CO_2) deriving from fossil fuels have gained important attention since establishing the Kyoto Protocol in 2005. Various technological approaches for CO_2 capture and reduction into added-value products have been proposed. Concurrently, cleaning the air from other pollutants such as ammonia, carbon monoxide, sulfur dioxide, and nitrous oxide is another major problem requiring the development of new technological solutions with extreme urgency.

4.1 Nanomaterials for CO_2 Capture and Reduction

Oxide-based nanostructures have been widely investigated as catalysts for CO_2 capture and reduction. For example, mesoporous silica materials captured considerable attention for several application fields, including adsorption and catalysis. Mesoporous silica materials show large pore volumes, tunable pore size, high surface areas, and rich surface chemistry [39, 40]. Thanks to the properties mentioned above, CO_2 and H_2S have been removed selectively from natural gas using amine-functionalized silica xerogels and ordered mesoporous silica [41]. At room temperature, 80% of the overall CO_2 amount (i.e., 50 mg/g of sorbent) was removed within 30 min, and 80% H_2S adsorption was obtained within 35 min. CO_2 capture and reduction into added-value products have also been achieved by exploiting other metals and metal oxide nanomaterials' catalytic properties. Cu- and CuO-based nanoparticles are by far the most promising ones. In this case, the conversion of CO_2 into hydrocarbons and alcohols has been demonstrated [42]. Similarly, noble metals such as Pt and Au NPs also show similar conversion properties. Due to their low-cost and Earth abundance, metal oxides such as MnO_x, TiO_2, and ZnO nanomaterials are currently under strong investigation for CO_2 catalytic reduction processes, and promising results in the reduction of CO_2 into non-toxic and added-value products have been recently obtained [42].

4.2 Air/soil Purification from Other Pollutants

CNTs have been recently proposed as adsorbent nanomaterials to remove toxic species from the air [43]. The adsorption capability of pollutants by CNTs derives from the very high surface area and porous structure of the CNTs themselves, as well as by the existence of various surface functional groups, both native or induced by specific chemical/thermal treatments, which make organic and inorganic toxic compounds prone to be anchored to the CNT surface. For example, the adsorption and removal of dioxin from a gaseous environment were possible thanks to a good chemical affinity between benzene rings of dioxin and CNTs. Activated carbon nanofibers have been proposed for the successful removal of NO species deriving from fossil fuel combustion as well. NO can be efficiently adsorbed on activated carbon's external surface because of the interaction taking place among specific functional groups. In other cases, CNTs have been used as adsorbent nanomaterials for the capture of NO_x and SO_2. Strikingly, the amount of NO_x absorption per gram of CNTs was approximately 78 mg [43].

Polymers are successfully used for the removal of contaminants such as heavy metals, gases (CO, SO_2, NO_x), and organic pollutants (hydrocarbons, drugs, or volatile organic compounds) [43]. For example, polymer nanoparticles are emerging as a new solution for soil remediation by exploiting the principle of surfactant micelles. These can show an amphiphilic behavior, being characterized by an outermost hydrophobic part and an inner hydrophilic one. Polyurethane nanoparticles have been engineered explicitly for the bioremediation of hydrophobic organic contaminants (phenanthrene) [44]. Polymers are also combined with inorganic nanomaterials to make hybrid materials. These have been studied for applications related to the environment, demonstrating good perspectives for removing several toxic metal ions and dyes.

5 Challenge N. 3: Health for Everyone

5.1 Tissue Regeneration

Nowadays, tissue regeneration is one of the biggest world challenges. It consists of creating newly synthesized tissues to replace damaged organs or other biological entities. The implementation of new tissue in human body is an important issue: before this step, the effects of such a procedure must be verified both in vitro and on test animals. Nowadays, one of the most powerful tools that can be used in this field is represented by *Organ-on-a-Chip* (OoC). These devices can have dual usage:

i. They can replace animal models in laboratory tests still obtaining accurate results: animal tissues are quite different from human ones, while OoCs can be designed to mimic human behavior; and
ii. They can be used as synthetic surrogates of biological tissues: this is the most challenging aspect since, while a human tissue can be reproduced with an OoC, its implantation is not so simple. Furthermore, it is not yet available a fast and accurate way of producing enough material to be used for this purpose. The industrial field's main efforts aim to increase the throughput of these devices using robotics, sensors, and advanced data processing methods [45].

At present, the main source of new tissues are human donors (autografts or allografts) or sometimes animal donors (xenografts), but these solutions cannot satisfy the global world's needs. One industrially exploited solution is SkinGun™ [46], a device that sprays stem cells' solution on damaged epithelial tissue to help its regeneration. This therapy allows treating a patient less than a week after skin burn, while conventional methods require several weeks for a complete recovery of all the damaged tissue.

However, it must be noted that this kind of therapy can have dangerous drawbacks: it is difficult to control stem cells' growth, and their use can lead to skin cancer or other diseases. Since SkinGun™ and other similar devices still need more experimentation to provide efficient human cell-based models, more investigation of these devices and models can predict and simulate human physiology.

Much scientific evidence exists about tissue regeneration methods for developing organoids, i.e., three-dimensional (3D) models, and on the study of microenvironments to direct stem cell differentiation [47]. Physiologically, clusters of cells produce and secrete the extracellular matrix (ECM), which consists of a supporting structure. The ECM also allows the transfer of various signaling biomolecules. The thorough knowledge of the various mechanism underlying cell behavior in response to stimuli from the environment and biomolecular signals, as well as their ability to cooperate in creating tissues and organs, allow not only a full comprehension but also the possibility to manipulate all these processes with synthetic scaffolds to heal damaged tissues or produce new ones [48].

The process typically starts by constructing a nano-micro material-based scaffold. Various sources are available, from proteins to natural or artificial polymers. Then, the cells can be incubated using a mixture of growth factors in the cell medium and, if proper culturing conditions are set, cells can proliferate and generate new tissue. Sometimes, the cells, scaffolds, and growth factors are first mixed to obtain a natural self-assembled tissue. However, synthetic ECM still shows some challenges, i.e., low efficiencies in cell proliferation and growth. Indeed, poor nutrition exchange in synthetic ECM is often present, impairing cell growth and survival. Microgels are microscopic hydrogels characterized by a large surface-to-volume ratio. Together with cells, microgel-based scaffolds can produce efficient mass transport and enhanced cell–matrix exchanges. Microgels can also be produced using microfluidic technology, obtaining high-efficient mass transport and spatially-controlled physicochemical properties, and obtaining building blocks that can be usefully assembled to build the final desired tissue.

Different tissue regeneration techniques involving nanostructures and nanomaterials are already exploited in both research and clinical applications [49]. Such nanomaterials have to sustain the following requirements:

I. biocompatibility, in terms of biological non-immunogenicity, low or null cytotoxicity, and absence of scar tissue formation, as well as mechanical biocompatibility (withstand hydrostatic pressure and maintain spacing);
II. biodegradability and bioresorbability, with a degradation rate that matches the cell or the tissue growth both in in vitro and in vivo *studies*; and
III. sterilizability to avoid infections without interfering with bioactivities and chemical composition.

Bone tissue engineering, for instance, can be achieved by assembling tissue-derived bone cells and organic/inorganic scaffolds. A scaffold is a macroscopic load-bearing structure, typically with a millimetric size or even more, which shows a micrometer size porosity. This allows cells to proliferate within them, and the presence of nanostructured pores increases the surface available for biomolecule hosting and further delivery [50, 51]. The porous scaffold still represents a supporting element for the damaged tissue. It supports the growth of new cells (starting from the patient's ones), allowing the diffusion of oxygen and nutrients.

The nanostructure's high surface area and properties can affect cell survival, proliferation, signaling, growth, and differentiation. All these factors are responsible for the final tissue healing success or failure, requiring urgent intervention at the clinical level.

Tissue Nanotransfection (TNT) is a new technology developed to repair or restore injured or aging tissue, including blood vessels, nerve cells, and entire organs [52]. It is composed of two constituents: a nanotechnological chip able to release the biological cargo to the skin cell and the biological cargo molecule itself, which enable the conversion of cells to a new function. This nano-based chip is a silicon wafer micro-patterned with a nanochannels array (having ten μm depth and around 500 nm width) within large micro reservoirs. The chip is applied topically to obtain the direct release of the cargo to the skin, but also dedicated electrodes can perforate the skin to deliver intradermally specific reprogramming factors, including DNA or proteins (biological cargo). A low-intensity electrical current is then applied to the chip's nanochannels to allow the nano-poration of the exposed cell membranes and allow factors reprogramming into cells. The large surface area of the nanochannels improves the interface between the cells and the chip.

For this reason, the process ends in few milliseconds. This technology was already able to restore the blood flow in mice models in vivo with injured legs and show animals' skin cells reprogramming to become vascular cells. New brain tissue from the animals' skin transplanted into the brain was also induced to recovery a stroke.

5.2 New Solutions for Replacing Antibiotics

The development of antibiotics has revolutionized the approach to treating and preventing infectious diseases, allowing modern medicine's evolution. However, the emergence of antibiotic resistance is likely to make these achievements futile. Whenever an antibiotic is administered, only resistant bacteria survive and become dominant, being the only ones capable of continuing to grow [53].

Antibiotic resistance is one of the main public health problems worldwide with important implications both from a clinical point of view (increased morbidity, lethality, disease duration, the possibility of developing complications, the possibility of epidemics) and in terms of economic impact. The problem of antibiotic resistance is complex since it recognizes several causes: the increased use of these drugs (including inappropriate use) in both human and veterinary medicine; the spread of infections caused by antibiotic-resistant microorganisms, typically in hospitals, accompanied by an inadequate control over the spread and nature of these infections [53].

The conventional practice to overcome the resistance to antibiotics is to avoid their use in general. In case of real need, the conventional practice restricts their use to cases where it is strictly necessary. Another possible solution is to exploit the antibacterial properties of silver. Therefore, plasters, gauze, or creams containing silver ions for cutaneous use are currently on the market [54]. Silver ions can penetrate bacterial membranes and destroy them by interacting with enzymes and proteins, causing deformation, loss of mobility, and inhibiting the multiplication processes. This method at the macroscale allows the treatment of exclusively superficial infections, being effective and non-invasive. However, it opens a plethora of possibilities when applied at the nanoscale. Most metal and metal oxide nanoparticles exhibit bactericidal properties by releasing metal ions and ROS generation, damaging the bacteria membrane, DNA, and many other macromolecules and bacterial cell functions. In particular, ROS affects the membrane's permeability, causing a breakdown of the bacterial cell's physiological balance that causes death. The research is focused on silver nanoparticles to exploit their nanometric properties, such as the high surface-volume ratio and their consequent high reactivity [55]. Ag^+ ions bind to thiol groups present in proteins and alter the cell membrane structure, neutralizing enzyme active sites and interfering in DNA for correct cell reproduction. Silver nanoparticles with a diameter lower than 10 nm can efficiently penetrate the cell membrane of bacteria and inhibit metabolic activities. For this reason, the smallest the particles, the greatest is the efficacy. Despite this, the studies are not completely exhaustive and more research is needed to deepen such findings.

Metal oxide nanoparticles were also proposed as effective against various bacteria (both Gram-positive and Gram-negative ones), viral species, and parasites. In particular, ZnO nanocrystals appeared as efficient nanoantibiotics [36]. Their efficacy might be attributable to various possible mechanisms:

I. The generation of ROS, due to the presence of ZnO crystalline defects;
II. The release of zinc cations (Zn^{2+}) altering the surface charge on the bacterial cell wall, leading to breakage; and
III. Electrostatic interactions between ZnO and microorganism cell membrane, having opposite surface charges. This interaction can act on the cellular metabolism, causing the death of the bacterium.

Regarding the applications, ZnO is already used in the food industry as a bactericide in packaging, while a variety of medicaments containing ZnO are currently produced in the pharmaceutical industry in the form of creams, lotions, and powders [56]. ZnO can also be photostimulated using ultraviolet light to produce ROS for antimicrobial applications, with mechanisms detailed in Challenge 2.

5.3 Fight Cancer Disease—New Diagnostics for Early Detection of Cancer or Circulating Tumor Cells

Cancer is the second death cause worldwide after cardiovascular diseases. Unfortunately, recent predictions estimate an increase of approximately 70% of new cases in the next years. Traditional anti-cancer therapies (surgery, chemotherapy, radiotherapies, or their combinations) have a certain degree of success and many disadvantages, including low specificity, causing side effects to the patients, or multidrug resistance developed in tumors.

Among the most innovative therapeutic and diagnostic strategies against cancer, nanomedicine has recently received great interest in the last decade. It is defined as "*the application of nanomaterials for health and medicine, and it is part of nanotechnology medical applications*" [57]. Nano-objects are eligible for nanomedicine treatments thanks to their small size, which allows exploiting the Enhanced Permeation and Retention effect (EPR), i.e., the accumulation into the tumor mass thanks to the fenestrations present in tumor-dedicated blood vessels. With these mechanisms, a passive targeting to cancer is achieved, reducing, from the other side, possible negative effects associated with a non-site selective administration of therapeutics [58]. Nanomaterials can be designed to carry and deliver one or more therapeutic agents and work as therapeutic agents themselves (i.e., upon stimulation or their intrinsic toxicity) or even used for diagnosis [59, 60]. Furthermore, to improve their cancer-targeting abilities, specific functionalization at nanomaterials surface can be used to attach targeting ligands [61]. Many therapeutic nanoparticles as both drug carriers or stimuli-responsive nanomaterials are currently under evaluation in clinical trials, and some of them have already been approved for clinical use [62, 63]. At present, the most used ones at the clinical level are nanoparticles as carriers of drugs. This strategy enables to improve either the therapeutic efficacy of the drug, or just to protect it from the biological environment, improving the drug bioavailability, or even to reduce drug clearance or related side effects. The first example of clinically approved and commercially available nanomedicine [64] consists of both liposomes entrapping chemotherapeutic, as Doxil (liposomes carrying doxorubicin drug), or protein aggregates, for example, Abraxane (an albumin particle containing the drug Paclitaxel). In recent years, some other types of nanopharmaceuticals formulations have been proposed against cancer, i.e., polymeric nanoparticles, in particular micelles, and different types of both organic and nano-formulations. In some instances, such nanoparticles can carry drugs and nucleic acids or target ligands specific to cancer cell receptors, thus opening a broad plethora of possible new solutions to treat cancer. In the last decade it was also proposed a theranostic approach, yet combining both therapeutic and diagnostic properties on nanoparticles and thus opening the research in many multidisciplinary fields [65].

Despite these great and exciting perspectives in nanomaterials for medicine, there is less understanding of their practical efficacy. Many findings should still be collected to demonstrate a real improvement concerning conventional treatments. Patient compliance, fast recovery rate, or decrease or suppression of side effects are still questionable when using nanomedicines. A possible reason can lie in the validity and prediction behavior of in vivo tests toward the final clinical validation due to interspecies differences between mice and humans [66]. Organ-on-chips have also been proposed as alternative methods to in vivo tests or reduce and refine numbers of animals and tests in vivo. However, they have not yet been established as a standard praxis, and still, open questions remain. At present, the gold standard resides in the careful design of nanoparticle parameters, including strict control of their synthesis process (to be controllable and highly reproducible), on the chemistry in particular at the surface (surface charge, hydrodynamic diameter, colloidal stability, biodegradation or reaction in biological media, protein corona formation, the tendency of aggregation), and of their physical, morphological and structural properties, (i.e., size and shape), and decoration with organic stealth materials and biomolecules for better stabilization in biological media and appropriate cell targeting. These design parameters will impact the nanoparticles' final pharmacokinetics and pharmacodynamic behavior on the toxicity profile and therapeutic efficacy [67].

In the field of nanotools used for diagnosis, the main problem to be faced is the early detection of cancer cells upon primary cancer establishment or as circulating tumor cells (CTC) once upon cancer relapses or metastatic development. In recent years, important progress has emerged in detecting CTC or tumor DNA (ctDNA) in the bloodstream and using one or both of them as tests for cancer screening. However, the current CTCs detection methodologies show some limitations to this early diagnosis approach as the bloodstream's CTC concentration is very low (one cell per million leukocytes and billions of erythrocytes). CTCs are highly heterogeneous (showing different morphologies and/or gene expression). Their detection is often facilitated when they adhere to tissues, thus already creating metastases. An efficient approach is using portable microdevices based on Parylene filters that allow the isolation of CTCs derived from human peripheral blood according to size [68]. This method also introduces a good selectivity for the size of the CTCs, which also allows characterizing their tumor origin.

Another possible approach currently under study and clinical trials is the use of nanomaterials for CTC detection [69]. A key feature of nanomaterials is represented by the considerably high exposed surface area, which allows efficient bonds with ligands useful to recognize cancer indicators (e.g., immuno-identifiers). Furthermore, to overcome CTC heterogeneity, the nano-roughness of materials can exploit strong cancer cells' adhesion preference compared to normal blood cells.

For this purpose, a NanoVelcro device was developed [70]: a system composed of four different chips containing different nanomaterials able, with a mechanism similar to that of the velcro, to hook, count, isolate and purify CTCs from the blood sample. The final purpose of NanoVelcro is not only for the efficient and early diagnosis but also for the study of new anti-cancer therapy methods. The structure is based on silicon nanowires (SiNS), possibly enriched with TiO_2 nanowires, gold clusters, or Fe_3O_4 nanoparticles, which, arranged vertically, create the "velcro" effect able to capture cancer cells with great efficiency (i.e., between 40 and 70%), much greater than a smooth Si substrate. To date, studies have been carried out on pancreatic, prostate, and kidney cancers, all with excellent results, which confirm the great prospects that this field of research offers and the importance of the development of this diagnostic trajectory as liquid biopsy.

Another system recently developed for identifying CTCs is the CellSearch™ device [71], highly enriched in immunomagnetic applications, fluorescent labeling, and rare cell population detection. It was introduced in 2004 as the first and only FDA-approved method designed for the enumeration of cancer cells circulating in only 7.5 ml of blood. It is also a liquid biopsy methodology where a specific blood sample is inserted into the device after being centrifuged to separate the solid components from blood plasma. CellSearch detects the presence of CTCs through ferrofluid magnetic nanoparticles, covered with special antibodies that target epithelial adhesion molecules. The targeted and thus dye-labeled cells are then collected by magnetic separation and located to a defined focal depth for optical analysis. Once scanned, the system shows the possible candidates for being positive cancer cells and presented them to an operator for final review.

To date, nanoparticles can be applied to the tumor diagnosis by conventional methods to enhance these techniques' sensitivity and specificity, thus working as contrast agents. For example, ultrasmall paramagnetic iron oxide (USPIO) nanoparticles and other magnetic and super-paramagnetic nanoparticles (for example, cased on manganese oxide) can be adopted as a contrast agent for magnetic resonance imaging (MRI). Furthermore, their surface functionalization with targeting ligands able to selectively bind to tumor tissue can easily evidence its presence.

Gold nanoparticles are also employed as a contrast agent in photoacoustic imaging, while recently, zinc oxide nanocrystals were proposed as echographic contrast agents [72].

Quantum dots were proposed as excellent photoexcitable optical imaging agents for tumor diagnosis [73]. QDs exploit quantum effects to emit radiation in a precise light spectrum thanks to quantum confinement, which depends on the particle size. Thus QDs, thanks to their precise optical emission, which provide a chromatic contrast of greater prominence than organic dyes, represent a possible improvement in the detection of cancer cells. In particular, they can be linked to fragments of antibodies binding to specific antigens present on cancer cells and, thanks precisely to the optical properties of QD, make them visible. Good results have been obtained regarding detecting breast, lung, and pancreas tumors [73].

6 Challenge N. 4 Restoration and Conservation of Public Heritage

The world's cultural heritage undoubtedly represents the sum of the incredibly vast pool of ideas that mankind has developed all along with its history. For this reason, it should be considered the starting point for each future social, economic, and technological advancement. Therefore, its preservation is of paramount importance to avoid losing the whole experience that our ancestors collected.

Buildings, monuments, paintings, and artistic products with an anthropological or historical significance can be subjected to aging. Indeed, external factors, like humidity, light, and material oxidation due to air exposure or harsh weather conditions, may modify the artifact's original appearance. In addition to these natural phenomena, some of today's impacting factors have also to be considered. The increasing pollution in our cities is boosting the aging of monuments, inducing irremediable damages to them.

If, on the one hand, public heritage provides an unfailing prerequisite for technological progress, on the other hand, scientific advancements may represent the key for cultural heritage preservation. In this sense, nanotechnologies are one of the most successful pieces of evidence to this statement. Indeed, the new physicochemical properties deriving from the reduced dimension of nanomaterials are an enormous advantage toward materials versatility that allows for effective protection against a wide range of potential art damages or an efficient instrument to restore the original conditions of artifacts.

6.1 Protective Films and Materials for Conservative Restoration

Wind, rain, extreme cold, or heat are all factors that contribute to the aging of all the monuments which are located outdoor, like statues or buildings. Indeed, monuments typically exposed to weathering are composed mainly of carbonates (for example, marble, lime), particularly susceptible to moisture and pollutant powder deposition.

In recent years, the urgency for finding a way to protect these art pieces has been raised because of climate changes. The presence of gases and rains rich in sulfur oxides is, in fact, one of the main causes of the stone surface recession [74] because of the reaction between SO_2 and carbonates that leads to the soluble calcium sulfate. Moreover, nitrogen and carbon oxides have to be added to the list of detrimental chemical species, with the latter responsible for the blackening of the artifact.

The aging process becomes relevant according to the nature of the stone composing the artifact [75]. For example, the presence of moisture in the stone influences in a major way the amount of absorbed acidic gases, and the porosity of the stone heavily affects this aspect.

Building maintenance is therefore required periodically on each monument, even if with a variable frequency. In some cases, this may lead to very high costs for both the intervention's difficulty and the high risk of ruining the artifact.

In the view of the restoration of the calcium-based building, calcium hydroxide has often been used because it is highly compatible with the stone's original chemical composition. CaOH can carbonate in the presence of carbon dioxide, generating the main component of building and wall paintings, i.e., calcium carbonate [76]. However, bulk CaOH is poorly soluble in water and cannot penetrate all the stone pores. Thanks to nanotechnology advancements, alcohol-stable CaOH nanoparticles could be prepared and allowed to fill the stone's finest pores and get complete carbonation of the system due to the higher reactivity of the nanoparticle.

From what concerns the protection perspective, nanomaterials have often been exploited. For example, TiO_2 nanoparticles are excellent photocatalytic semiconductors used to generate protective and self-cleaning films on monuments [77]. As described above, light-activated TiO_2 can generate radical species, ROS, which may, in turn, degrade various pollutants, then convert them into less harmful molecules.

Also, nanotechnologies may provide an efficient way to limit weathering's detrimental effects by generating hydrophobic surfaces. In particular, the exploitation of silica (SiO_2), alumina (Al_2O_3), or tin oxide (SnO_2) nanoparticles embedded in silicon rubber matrices is gaining much attention. The reason lies in the ease of realization of the composite and the effectiveness of generating hydrophobicity because of a coupled effect between surface roughness and surface tension modification.

Silicon-based materials are extensively exploited to strengthen the porous matrix of the stone [78]. In fact, like tetra-ethyl ortho-silicate (TEOS), alkoxysilane products are often used as a filler for the stones' pores because their very low surface tension allows deep penetration of the materials into the object to be recovered. TEOS's polymerization into a microporous structure protects the stone from moisture, but it may be responsible for some cracks because of the pressure released during drying. The inclusion of some oxide nanoparticles can increase the pores' dimension and toughen the material to solve this problem. The challenge with these systems is that these particles' inclusion leads to a not perfectly transparent system, which may affect the art piece's legibility.

A good compromise has been found with composite materials made of poly-siloxane and silica nanoparticles of dimensions of about 14 nm [79]. The new material was tested on three different stones (limestone, sandstone, and granite), showing a complete change of the surface behavior from hydrophilic to hydrophobic. Moreover, this material did not significantly affect the treated stone's appearance, differently from other more common techniques.

6.2 Protection and Restoration of Paintings

Both on walls and canvas, paintings represent a large portion of our cultural heritage, perhaps being the most important visual art that reports visual testimonies of our ancestors' life and beliefs. However, these art pieces include the coexistence on the surface of a huge amount of materials, and their cleaning requires an extremely high selectivity to avoid the elimination of part of the painting. Moreover, the agents that may affect painting legibility range from inorganic modifications to microorganisms, and therefore, a single remedy for all these aspects is quite difficult to find.

In the past, acrylic and vinyl polymers were used to perform the restoration. However, their use has been shown to induce painting deterioration because their removal requires aggressive solvents that are poorly selective to the artifact's outer surface [80]. Also, but not less important, pure organic solvents used for cleaning are often toxic, also leading to danger for the operator.

An advancement in this field is provided by micro and nano featured fluids like microemulsion. Microemulsions are systems composed of two insoluble liquids where one of them is dispersed into the other. The most exploited one in the restoration of cultural heritage is oil in water emulsions, where oil is dispersed into the water thanks to amphiphilic molecules' help. In fact, amphiphilic molecules can orient themselves to expose their hydrophilic tail toward the water and their hydrophobic part toward the insoluble oil. In this way, thermodynamically stable microbubbles of oil are formed.

The advantages of these systems are manifold. First of all, the amount of solvent inside the cleaning solution is heavily reduced, leading to lower toxicity. Moreover, the system's selectivity is increased since the solubilization of the dirt is limited in space and can be further controlled, as well as a major reactivity is obtained due to the higher exposed solvent surface. Finally, the aqueous solution does not require any further apolar solvent to be removed, avoiding the risk of dissolving actual painting layers.

Sometimes, instead of using microemulsion systems, gel matrices are loaded with commonly used cleaning solvents to control their release in the deep pores of the painting and to reduce the detrimental phenomenon that the solvent may induce in its pure form. However, removing these gels from the painting may represent a critical task on very fragile surfaces because it typically requires further solvents, which may remain stuck on the work of art. Nanomagnetic particles embedded in gels have been proposed to solve this problem [81]. In particular, nanomagnetic beads were embedded in a polyethylene glycol (PEG)—acrylamide gel to form a magnetic stimulation active-gel. The painting surface can consequently eliminate without leaving any residual.

7 Challenge N.5 Reduced Impact in Human Mobility

7.1 Improved Aerospace Vehicles

In 1903 the Wright brothers and their Flyer made our world smaller. The airplane's invention can be considered one of the most important military stones of the twentieth century, representing a revolution of our lifestyle. Many technological improvements have been made since that date, and now we can travel from one continent to another in a few hours. However, one of the most astonishing achievements is the possibility to leave our planet, which is acquiring more and more interest in recent years, not only because of scientific, economic, and military purposes but even in terms of space tourism. After the moon landing in 1969, the next goal is to land on Mars, the first planet where men have posed their feet after Earth.

Such an ambitious destination suffers from many challenges that have to be overcome, such as the effect that this journey may have on the travelers and safety-related issues, the amount of energy required for this kind of trip, and the huge costs, that is probably the main issue impacting on the development of this field.

As may be easily guessed, modern material science with a particular role of nanotechnology may partially solve some of these problems by optimizing more and more the shuttles which are supposed to bring us far from home. Highly insulating materials, for example, aerogels, may protect the electronics of the shuttles from the very highly harsh environmental conditions present in space. New energy harvesting and storage devices, like the one presented in the previous sections, may be a useful alternative for a source of electric energy after the arrival of the shuttle in space. Finally, carbon-based materials may improve the stability of the structural parts of the spacecraft, lowering at the same time their weight and leading to an improvement in the propulsion efficiency.

In this brief section, some of the nanomaterials' possible applications will be listed to highlight how this field of science may improve space travel performances.

7.2 Aerogel for Aerospace Applications

Aerogels are nanostructured systems composed of a matrix full of nanopores. This peculiar structure gathers many interesting properties, like extremely low density and high surface area, very poor thermal conductivity, transparency, acoustic insulation, and low dielectric constant.

The preparation of aerogels was made possible only by forcing the substitution of the liquid part of a gel with gas without the collapse of the structure. This is possible only by inducing the liquid's evaporation, autoclaving the system in high pressure and temperature. The first successful attempt was made in 1931 by Kistler [82] when he managed to obtain the first silica aerogel. From that date, several potential applications of aerogels in the aerospace field have been proposed.

Most of its applications are currently used in scientific research. Silica aerogels have often been used as hypervelocity particle capture systems. Indeed, the physics of particles colliding at a speed higher than one km/s typically involves the projectile's vaporization that cannot be therefore directly studied. However, the very low density of silica aerogels (~ 0.1 g/cm^3) allows the particle traveling at such a high speed to hit the solid sample without damages that prevent their analysis and simplified the more complex, usually exploited indirect procedures [83].

Aerogels' low thermal conductivity has triggered the idea of using these materials to protect electronics or delicate parts of both shuttles and satellites from the very harsh temperature conditions that may be found in space. For example, aerogels were successfully used to protect the battery pack of an X-ray spectrometer of a Mars Rover in the Pathfinder mission of 1997. These materials have also been used to produce space suits that are incredibly lighter and thinner concerning the usually exploited ones because of the low density of aerogel and their thermal insulation efficiency [84].

Despite the still-high costs of aerogels, industrial applications' advantage is related to the energy and money savings that efficient thermal insulation may provide in all these systems. In fact, the exploitation of these materials would prevent the use of active thermal devices that require fuel.

7.3 Carbon Nanotubes in Aeronautics

Apart from smart designs aimed at improving the aerodynamics of aircraft and optimizing their engines' efficiency, aeronautics and aerospace costs are reduced through the decrease of planes' weight. Airplanes are devices that typically have a life of some tens of years, and a reduction of their weight may mean a long-term energy saving in terms of propellant.

Research has faced this challenge by synthesizing novel nanomaterials able to substitute the metallic parts of the plane. The most promising results are achieved with carbon nanotubes (CNTs) (Fig. 3). CNTs are self-rolled graphene planes, which results in a cave cylinder with a single atom surface thickness. They may be composed of a single nanotube (single-wall carbon nanotubes) or several concentric

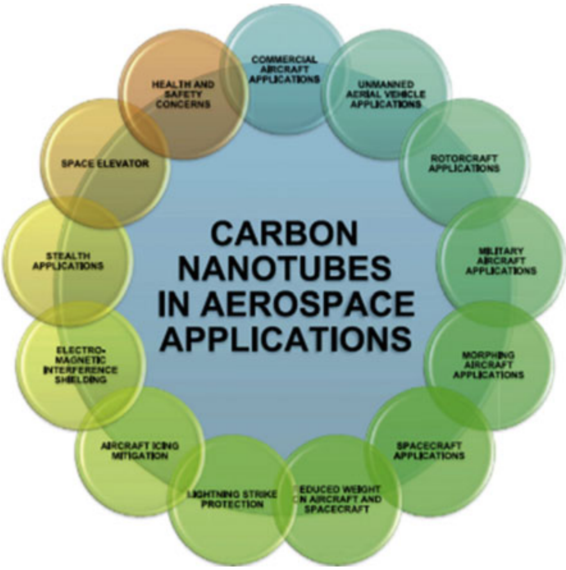


Fig. 3 Overview of CNTs applications for aeronautics and aerospace. (Adapted with permission from [85])

nanotubes (multi-wall carbon nanotubes). The peculiar configuration of the atoms in CNTs provides fascinating mechanical, electrical, and thermal properties, which may positively affect airplanes' performances and costs [85].

From the structural point of view, polymeric nanocomposites, i.e., polymeric matrices filled with nanostructured materials, may represent an interesting alternative to substitute structural parts. Indeed, nanocomposites, in which the nanostructured filler is a carbon-based system, are already used in the market, and the use of nanotubes would increase even more the aircraft's performances. The reason is that Carbon Nanotubes Reinforced Polymers (CNRP) not only present an incredibly lower density when compared to usually used alloys but also improved performance in terms of flexibility and structural strength. In summary, reduced weight is brought without damaging the structural integrity of the vehicle.

The good electrical conductivity of CNTs also suggests replacing the wiring system that connects the whole airplane electronically. The mole of copper wires present in modern airplanes has a big role in the vehicle's total weight, and its substitution with electrically conductive cables based on carbon nanotubes would be a step forward in an ultra-light airplane [85].

Moreover, the electronic properties of CNTs make them suitable to passively collect data about surface deformation, which could be used to sense eventual damages and increase the device's safety [85].

CNTs may be exploited in many fields of aeronautics, making them the most used and promising nanostructure in this field. Applications of CNTs against aircraft icing or to develop stealth military vehicles [86] are present in literature and will be very likely present in the next generation of airplanes [87].

8 Conclusion

The role of nanotechnology in facing the current world's challenges is acquiring increasing importance. We have reported here a list of the biggest world's challenges. In our opinion, nanotechnology solutions can be relevant totally or in part, i.e., contributing with other technologies to this challenge's solution. Some of the possible solutions are already in place; others are under investigation at the research levels. Even others still require a lot of work and commitment by academic, industrial, and even policymakers to become reliable solutions. We hope that all the presented nanotechnology researches will become industrially viable solutions to the world's challenges and improve our quality of life, contributing to a better world.

Core Messages

- Can the greatest problems of our contemporary world be solved? Yes, and many inspirations and solutions, partially or entirely, can come from the nanoscience world.
- In this chapter, we analyzed the most recent potential solutions for climate change, clean, renewable energy, clean water supply, polluted soil, air and water remediation, health for everyone, and availability of life-saving pharmaceuticals.
- A broad, visionary, and multidisciplinary approach is needed and is in our hands with Nanotechnology's exploitation.

References

1. Bahar H (2019) Tracking power—solar PV. In
2. Green MA, Hishikawa Y, Dunlop ED, et al (2019) Solar cell efficiency tables (Version 53). 27:3–12
3. Gangopadhyay U, Jana S, Das S (2013) State of art of solar photovoltaic technology. Conf Papers Energy 2013:764132
4. Pse (2020) PV production by technology—percentage of global annual production. In: Frauhofer (ed), online, p https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/Photovoltaics-Report.pdf
5. O'regan B, Grätzel M (1991) A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. Nature 353:737–740
6. Le VH, Nguyen TH, Nguyen HH et al (2020) Fabrication and electrochemical behavior investigation of a Pt-loaded reduced graphene oxide composite (Pt@rGO) as a high-performance cathode for dye-sensitized solar cells. Int J Photoenergy 2020:8927124
7. Baruah J, Narayan J (2019) Mini review quantum dot sensitized solar cell: introduction
8. Wang ZL (2020) Triboelectric nanogenerator (TENG)—sparking an energy and sensor revolution. 10:2000137

9. Pu X, Li L, Song H, et al (2015) A self-charging power unit by integration of a textile triboelectric nanogenerator and a flexible lithium-ion battery for wearable electronics. 27:2472–2478
10. Rivera VF, Auras F, Motto P, et al (2013) Length-dependent charge generation from vertical arrays of high-aspect-ratio ZnO nanowires. Chem—Eur J n/a-n/a
11. Stassi S, Cauda V, Ottone C et al (2015) Flexible piezoelectric energy nanogenerator based on ZnO nanotubes hosted in a polycarbonate membrane. Nano Energy 13:474–481
12. Chiolerio A, Lombardi M, Guerriero A et al (2013) Effect of the fabrication method on the functional properties of BaTiO3: PVDF nanocomposites. J Mater Sci 48:6943–6951
13. Cauda V, Stassi S, Bejtka K et al (2013) Nanoconfinement: an effective way to enhance PVDF piezoelectric properties. ACS Appl Mater Interfaces 5:6430–6437
14. Cauda V, Torre B, Falqui A et al (2012) Confinement in oriented mesopores induces piezoelectric behavior of polymeric nanowires. Chem Mater 24:4215–4221
15. Cauda V, Stassi S, Lamberti A et al (2015) Leveraging ZnO morphologies in piezoelectric composites for mechanical energy harvesting. Nano Energy 18:212–221
16. Zhou G, Li F, Cheng H-M (2014) Progress in flexible lithium batteries and future prospects. Energy Environ Sci 7:1307–1338
17. Zhao Q, Zhu Q, Miao J, et al (2019) Flexible 3D porous MXene foam for high-performance lithium-ion batteries. 15:1904293
18. Yang P, Gao X, Tian X et al (2020) Upgrading traditional organic electrolytes toward future lithium metal batteries: a hierarchical nano-SiO_2-supported gel polymer electrolyte. ACS Energy Lett 5:1681–1688
19. Cheng Y, Xiao X, Pan K et al (2020) Development and application of self-healing materials in smart batteries and supercapacitors. Chem Eng J 380:122565
20. He H, Fu Y, Zhao T et al (2017) All-solid-state flexible self-charging power cell basing on piezo-electrolyte for harvesting/storing body-motion energy and powering wearable electronics. Nano Energy 39:590–600
21. Zhao K, Wang Y, Han L et al (2019) Nanogenerator-based self-charging energy storage devices. Nano-Micro Lett 11:19
22. Canavese G, Stassi S, Fallauto C et al (2014) Real-time pedobarography analysis by piezoresistive wearable insole. Sens Lett 12:1427–1432
23. Canavese G, Stassi S, Fallauto C et al (2014) Stretchable and wearable piezoresistive insole for continuous pressure monitoring. Key Eng Mater 605:474–477
24. Someya T (2013) Building bionic skin. IEEE Spectr 50:50–56
25. Yokota T, Zalar P, Kaltenbrunner M, et al (2016) Ultraflexible organic photonic skin. 2: e1501856
26. Lee B, Oh J-Y, Cho H et al (2020) Ultraflexible and transparent electroluminescent skin for real-time and super-resolution imaging of pressure distribution. Nat Commun 11:663
27. Ma L, Dong X, Chen M et al (2017) Fabrication and water treatment application of carbon nanotubes (CNTs)-based composite membranes: a review. Membranes 7:16
28. Surwade SP, Smirnov SN, Vlassiouk IV et al (2015) Water desalination using nanoporous single-layer graphene. Nat Nanotechnol 10:459–464
29. Fischbein MD, Drndić M (2008) Electron beam nanosculpting of suspended graphene sheets. 93:113107
30. O'hern SC, Stewart CA, Boutilier MSH, et al (2012) Selective molecular transport through intrinsic defects in a single layer of CVD graphene. ACS Nano 6:10130–10138
31. O'hern SC, Jang D, Bose S, et al (2015) Nanofiltration across defect-sealed nanoporous monolayer graphene. Nano Lett 15:3254–3260
32. Buelke C, Alshami A, Casler J et al (2018) Graphene oxide membranes for enhancing water purification in terrestrial and space-born applications: state of the art. Desalination 448:113–132
33. Fathizadeh M, Xu WL, Zhou F, et al (2017) Graphene oxide: a novel 2-dimensional material in membrane separation for water purification. 4:1600918

34. Hong X, Wen J, Xiong X et al (2016) Shape effect on the antibacterial activity of silver nanoparticles synthesized via a microwave-assisted method. Environ Sci Pollut Res Int 23:4489–4497
35. Raza MA, Kanwal Z, Rauf A, et al (2016) S. Size- and shape-dependent antibacterial studies of silver nanoparticles synthesized by wet chemical routes. Nanomaterials 6:74
36. Garino N, Sanvitale P, Dumontel B et al (2019) Zinc oxide nanocrystals as a nanoantibiotic and osteoinductive agent. RSC Adv 9:11312–11321
37. Damodaran VB, Bhatnagar D, Leszczak V et al (2015) Titania nanostructures: a biomedical perspective. RSC Adv 5:37149–37171
38. Laurenti M, Garino N, Canavese G et al (2020) Piezo- and photocatalytic activity of ferroelectric ZnO: Sb thin films for the efficient degradation of rhodamine-β dye pollutant. ACS Appl Mater Interfaces 12:25798–25808
39. Cauda V, Schlossbauer A, Kecht J et al (2009) Multiple core−shell functionalized colloidal mesoporous silica nanoparticles. J Am Chem Soc 131:11361–11370
40. Guerra FD, Attia MF, Whitehead DC et al (2018) Nanotechnology for environmental remediation: materials and applications. Molecules 23:1760
41. Huang HY, Yang RT, Chinn D et al (2003) Amine-grafted MCM-48 and silica xerogel as superior sorbents for acidic gas removal from natural gas. Ind Eng Chem Res 42:2427–2433
42. Mandal L, Yang KR, Motapothula MR et al (2018) Investigating the role of copper oxide in electrochemical CO_2 reduction in real time. ACS Appl Mater Interfaces 10:8574–8584
43. Yunus IS, Harwin KA et al (2012) Nanotechnologies in water and air pollution treatment. Environ Technol Rev 1:136–148
44. Tungittiplakorn W, Lion LW, Cohen C et al (2004) Engineered polymeric nanoparticles for soil remediation. Environ Sci Technol 38:1605–1610
45. Probst C, Schneider S, Loskill P (2018) High-throughput organ-on-a-chip systems: current status and remaining challenges. Curr Opin Biomed Eng 6:33–41
46. Renovacare https://www.renovacareinc.com/technology/. In
47. Jiang W, Li M, Chen Z et al (2016) Cell-laden microfluidic microgels for tissue regeneration. Lab Chip 16:4482–4506
48. Chen G, Kawazoe N (2018) Biomimetic extracellular matrices and scaffolds prepared from cultured cells. In: Springer (ed) Cutting-edge enabling technologies for regenerative medicine. Advances in experimental medicine and biology. Singapore
49. Dvir T, Timko BP, Kohane DS et al (2011) Nanotechnological strategies for engineering complex tissues. Nat Nanotechnol 6:13–22
50. Cauda V, Fiorilli S, Onida B et al (2008) SBA-15 ordered mesoporous silica inside a bioactive glass–ceramic scaffold for local drug delivery. J Mater Sci: Mater Med 19:3303–3310
51. Mortera R, Onida B, Fiorilli S et al (2008) Synthesis and characterization of MCM-41 spheres inside bioactive glass–ceramic scaffold. Chem Eng J 137:54–61
52. Miller MA (2017) Nanotransfection brings progress that's more than skin-deep. 9:eaao4216
53. Ventola CL (2015) The antibiotic resistance crisis: part 1: causes and threats. P T 40:277–283
54. Leaper D (2012) Appropriate use of silver dressings in wounds: International consensus document. 9:461–464
55. Arya G, Sharma N, Mankamna R, et al (2019) Antimicrobial silver nanoparticles: future of nanomaterials. In: Springer (ed) Microbial nanobionics. Cham
56. Racca L, Canta M, Dumontel B, et al (2018) Zinc oxide nanostructures in biomedicine. In: Ciofani G (ed) Smart nanoparticles for biomedicine. Elsevier, pp 171–187
57. Webster TJ (2006) Nanomedicine: what's in a definition? Int J Nanomedicine 1:115–116
58. Liu Y, Sun D, Fan Q et al (2020) The enhanced permeability and retention effect based nanomedicine at the site of injury. Nano Res 13:564–569
59. Cauda V, Engelke H, Sauer A et al (2010) Colchicine-loaded lipid bilayer-coated 50 nm mesoporous nanoparticles efficiently induce microtubule depolymerization upon cell uptake. Nano Lett 10:2484–2492

60. Dumontel B, Susa F, Limongi T et al (2019) ZnO nanocrystals shuttled by extracellular vesicles as effective Trojan nano-horses against cancer cells. Nanomedicine 14:2815–2833
61. Barui S, Cauda V (2020) Multimodal decorations of mesoporous silica nanoparticles for improved cancer therapy. Pharmaceutics 12
62. Bobo D, Robinson KJ, Islam J et al (2016) Nanoparticle-based medicines: a review of FDA-approved materials and clinical trials to date. Pharm Res 33:2373–2387
63. Racca L, Cauda V (2020) Remotely activated nanoparticles for anticancer therapy. Nano-Micro Lett 13:11
64. Weissig V, Pettinger T, Murdock N (2014) Nanopharmaceuticals (part 1): products on the market. Int J Nanomedicine 9:4357–4373
65. Lim E-K, Kim T, Paik S et al (2015) Nanomaterials for theranostics: recent advances and future challenges. Chem Rev 115:327–394
66. Gruber FP, Hartung T (2004) Alternatives to animal experimentation in basic research. Altex 21:3–31
67. Canta M, Cauda V (2020) The investigation of the parameters affecting the ZnO nanoparticle cytotoxicity behaviour: a tutorial review. Biomater Sci 8:6157–6174
68. Lin HK, Zheng S, Williams AJ, et al (2010) Portable filter-based microdevice for detection and characterization of circulating tumor cells. 16:5011–5018
69. Huang Q, Wang Y, Chen X et al (2018) Nanotechnology-based strategies for early cancer diagnosis using circulating tumor cells as a liquid biopsy. Nanotheranostics 2:21–41
70. Jan YJ, Chen J-F, Zhu Y et al (2018) NanoVelcro rare-cell assays for detection and characterization of circulating tumor cells. Adv Drug Deliv Rev 125:78–93
71. Andree KC, Van Dalum G, Terstappen LWMM (2016) Challenges in circulating tumor cell detection by the cell search system. Mol Oncol 10:395–407
72. Ancona A, Troia A, Garino N et al (2020) Leveraging re-chargeable nanobubbles on amine-functionalized ZnO nanocrystals for sustained ultrasound cavitation towards echographic imaging. Ultrason Sonochem 67:105132
73. Ramos-Gomes F, Bode J, Sukhanova A et al (2018) Single- and two-photon imaging of human micrometastases and disseminated tumour cells with conjugates of nanobodies and quantum dots. Sci Rep 8:4595
74. Bonazza A, Messina P, Sabbioni C et al (2009) Mapping the impact of climate change on surface recession of carbonate buildings in Europe. Sci Total Environ 407:2039–2050
75. Grossi CM, Murray M (1999) Characteristics of carbonate building stones that influence the dry deposition of acidic gases. Constr Build Mater 13:101–108
76. Baglioni P, Chelazzi D, Giorgi R et al (2014) Commercial Ca(OH)2 nanoparticles for the consolidation of immovable works of art. Appl Phys A 114:723–732
77. Licciulli A, Calia A, Lettieri M et al (2011) Photocatalytic TiO_2 coatings on limestone. J Sol-Gel Sci Technol 60:437–444
78. Mosquera MJ, De Los Santos DM, Montes A et al (2008) New nanomaterials for consolidating stone. Langmuir 24:2772–2778
79. De Ferri L, Lottici PP, Lorenzi A et al (2011) Study of silica nanoparticles—polysiloxane hydrophobic treatments for stone-based monument protection. J Cult Herit 12:356–363
80. Baglioni P, Berti D, Bonini M et al (2014) Micelle, microemulsions, and gels for the conservation of cultural heritage. Adv Coll Interface Sci 205:361–371
81. Bonini M, Lenz S, Giorgi R et al (2007) Nanomagnetic sponges for the cleaning of works of art. Langmuir 23:8681–8685
82. Kistler SS (1932) Coherent expanded-aerogels. J Phys Chem 36:52–64
83. Burchell MJ, SaJ F, Foster NJ et al (2009) Hypervelocity capture of particles in aerogel: dependence on aerogel properties. Planet Space Sci 57:58–70
84. Bheekhun N, Abu Talib AR, Hassan MR (2013) Aerogels in aerospace: an overview. Advan Mater Sci Eng 2013:406065

85. Gohardani O, Elola MC, Elizetxea C (2014) Potential and prospective implementation of carbon nanotubes on next generation aircraft and space vehicles: a review of current and expected applications in aerospace sciences. Prog Aerosp Sci 70:42–68
86. Ahmad H, Tariq A, Shehzad A, et al (2019) Stealth technology: methods and composite materials—a review. 40:4457–4472
87. Eseev M, Goshev A, Kapustin S et al (2019) Creation of superhydrophobic coatings based on MWCNTs Xerogel. Nanomaterials 9:1584

Marco Carofiglio graduated in Physical Engineering at the Politecnico di Torino in 2017 and got his M.Sc. degree in Nanotechnologies for ICTs in 2019 from the same university. During his Master's thesis work, he focused on developing ZnO-based piezoelectric smart scaffolds to stimulate biological tissues. He is currently attending a Ph.D. course in Chemical Engineering at Politecnico di Torino. Marco is working on the synthesis and characterization of doped ZnO nanoparticles and substrates for biomedical applications. His research work also includes the use of oxides for catalytic water splitting.

Valentina Cauda graduated in Chemical Engineering in 2004 at Politecnico di Torino and then received her Ph.D. in Material Science and Technology in 2008. After a short period at the University of Madrid, she worked as a post doc at the University of Munich, Germany, on nanoparticle drug delivery and tumor cell targeting. From 2010 to 2015, she worked as Senior Post Doc at the Istituto Italiano di Tecnologia in Torino. She is at present Associate Professor at Politecnico di Torino where she leads a group of 15 people. She holds two grants from the European Research Council (ERC Stg and ERC PoC) with the projects TrojaNanoHorse and XtraUS. For her research results, she received in 2010 the prize for young researchers at the Chemistry Department of the University of Munich. In 2013, she got the Giovedì Scienza award, in 2015 the Zonta Prize for Chemistry, and in 2017 the USERN Prize for Biological Sciences. Valentina is the author of 104 papers in peer-reviewed international journals (HF 34), and she holds four international patents.

9 The Kaleidoscopic Image as an Intersection Between Art and Science: From Microscopic Forms to Cosmological Models

Mari Nieves Vergara

From this I reach what I might call a philosophy; at any rate it is a constant idea of mine; that behind the cotton wool is hidden a pattern; that we –I mean all human beings– are connected with this; that the whole world is a work of art; that we are parts of the work of art. [...] Mine is that there is a pattern hid behind the cotton wool.

Virginia Woolf [1]

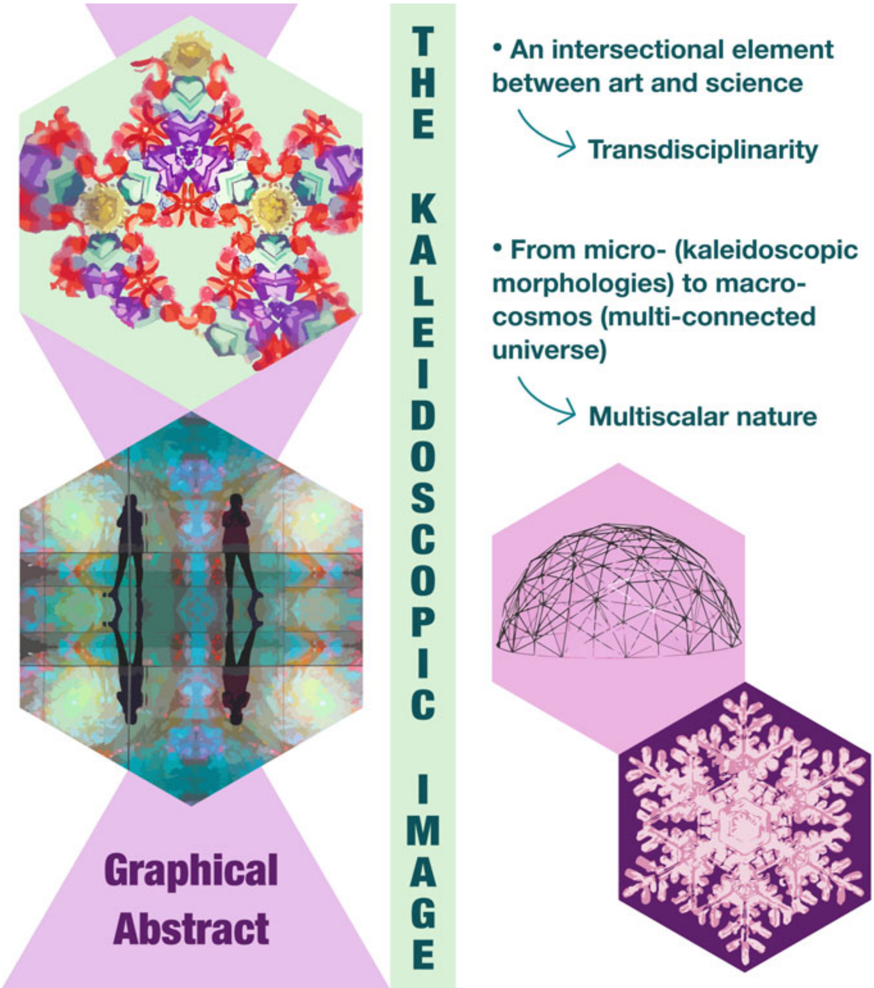
Summary

This chapter deepens on the structural connections and the visual correspondence between the morphology of different forms from nature and cosmological models or representations of the universe under the denomination of being kaleidoscopic. After a brief introduction on the kaleidoscopic issue from different aspects (Sect. "1"), its relation to fractal geometry, i.e., the image constructed by the kaleidoscope, will be considered an intersectional element between the fields of art and science. Within this platform, some cases will be presented, going from micro- to macro-cosmos. Microscopic forms such as diatoms, snowflakes, viruses, or fullerenes (Sects. "2", "3", and "4"), along with possible topologies and other representations of the universe (Sect. "5"), are categorized as kaleidoscopic. The connecting visual motif that will join all of these examples is a kaleidoscopic image. The comparison based on visual correspondence to study issues commonly related to scientific disciplines is an

M. N. Vergara (✉)
Department of Contemporary Art History, Fine Arts School, Complutense University of Madrid, C/ Pintor El Greco 2, 28040 Madrid, Spain

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_9

important aspect for artistic research, concretely for the visual studies area. This transdisciplinary approach includes recognizing particularly hybrid cases associated with kaleidoscopic configurations, like protein design. It concludes that the kaleidoscopic image manifests a multiscale nature (Sect. "6") since it is a configuration present from microscopic forms to the representations of a multi-connected universe.



Kaleidoscopic nature.

The code of this chapter is 01001011 01100001 01101100 01100101 01101001 01100100 01101111 01110011 01100011 01101111 01110000 01100101.

1 Introduction: The Kaleidoscopic Image as an Intersectional Element Between Art and Science

Virginia Woolf thought that behind the cotton wool, there is a hidden pattern, that everything is part of a connected work of art, which is the entire world [1]. Certainly, she *visualized the invisible* thanks to the creative intuition of recognizing new patterns, those artistic processes stated by the artist Agnes Denes in her *Manifesto*, as a way of questioning the accepted reality [2]. Surprisingly, Woolf's literary work resulted in being multifractal [3]. Accordingly, there is a hidden pattern *behind* her writing, which at the same time is related to the kaleidoscopic experience [4],[1] forming a correspondence between artistic work and scientific discovery.

In the Renaissance, art and science were highly related disciplines since individuals looked at knowledge as a whole. Nevertheless, from the seventeenth century, the cosmological view of knowledge has progressively partitioned into increasingly specialized disciplines [5]. This historical fact has provided more rigorous knowledge. However, the connection between disciplines and the findings that could have appeared from their interaction had been lost.[2] Nowadays, after some collaborative work between different areas within the last two centuries, a new paradigm in the relationship between art and science related to the kaleidoscopic nature is emerging after being essentially separated disciplines [6].

This chapter shows the research on some kaleidoscopic images from the morphology of microscopic forms to the structural and conceptual connections with multi-connected cosmological models and representations of the universe. In this regard, some cases will be presented, going from micro- to macro-cosmos. The study of kaleidoscopic images is of high importance for visual studies in artistic research since the kaleidoscopic quality appears to play an important role as an intersectional element between the fields of art and science. In this line, a recently published article [7] presented the results of a study on the kaleidoscopic condition of quantum behavior and the multiverse theory. It reflected on Borges' literary work and displayed an artistic research experiment reflecting on phenomena such as interference or the double-slit experiment was pointed out by Vergara [7]. By the end of this chapter, it will be noticed how the morphology in nature and other theories or representations from the scientific field are highly interrelated to the kaleidoscopic issue, which opens a transdisciplinary scope regarding the collaboration between the fields of art and science.

The kaleidoscope is an optical device invented in 1815 by physicist David Brewster. Although nowadays it is considered as a toy, its invention was aiming to be a medium with practical application in the work of designers and artists. In this way, it promoted an industrial paradigm that started at the beginning of the

[1] Many literary works that have been categorised as kaleidoscopic and usually inscribed within the stream of consciousness literary style—works such as the *Ulysses* (James Joyce) or *Hopscotch* (Julio Cortázar)—present a multifractal pattern as in Woolf's narrative [3,4].

[2] There are some cases in which art, science, and other disciplines were collaborating, although it was not the most common line to follow, almost not like in the Renaissance period.

nineteenth century, with the scientific purpose of revealing and democratize the techniques of the illusion [8, 9]. In Brewster's words, the kaleidoscope "*will create, in a single hour, what a thousand artists could not invent in the course of a year […] with a corresponding beauty and precision*" [10]. Previous to the kaleidoscope's invention, we find many geometrical decorative patterns from different cultures that present a kaleidoscopic condition. This device has influenced many artists, writers, and scientists from the nineteenth century to the present day. It has held a distinguished position with the first abstract representations at the beginning of the twentieth century [9], as well as due to its presence as a primary connective structure in Agnes Denes' work during the pre-internet era [11].

Nowadays, there are a wide variety of kaleidoscopes. The kaleidoscopic image can show colorful images and abstract shapes (Fig. 1a) when the tube has color pieces on the opposite side of the eye viewer hole. One can consider this kaleidoscope as a classic kaleidoscope. When it shows a kaleidoscopically fragmented reality, where mirrors reflect surrounding elements, it is called a telescopic kaleidoscope, also known as teleidoscope. When the term "kaleidoscopic image" is referred to, it could indicate a kaleidoscopic shape or the item's kaleidoscopic condition due to its similarity with this image. At the same time, it should be specified that the kaleidoscope would be the one constituted by a cylindrical tube with three mirrors inside the device, placed forming the shape of an equilateral triangle. Therefore, the structure that organizes the image through the kaleidoscope will be visually like an isometric grid that will remain denominated as a kaleidoscopic structure (Fig. 1b). This structure is invisible to the eye but remains as the structural component that organizes the image kaleidoscopically.

Throughout this chapter, the results presented mainly come from the kaleidoscopic cases' visual condition, except for Sect. "5.1", in which some theoretical issues are involved. Hence, most of the results will refer to the kaleidoscopic configuration in microscopic forms or cosmologies that belong to the scientific research field, aiming to categorize these different cases as "kaleidoscopic" depending on their visual correlation with the mentioned image. In some cases, also the creative motivation would be associated with the production of kaleidoscopic compositions, as will be seen in the kaleidoscopic arrangements that are presented in Sect. "2". Also, hybrid examples of transdisciplinarity between scientific research and art will be presented, such as the design of proteins or Buckminster Fuller's influence in discovering viruses' icosahedral structure in Sects. "3" and "4", respectively.

The high presence of kaleidoscopic configurations in nature is principally due to the connection between fractal geometry and the kaleidoscopic image. Since the view of kaleidoscopic images can be defined as a virtually infinite and repetitive geometric vision, it shares a visual correspondence with the geometry of nature, presenting a fractal condition. The most relevant properties of fractal geometry are also related to the kaleidoscopic image and structure. These properties are iteration —infinite repetition of an element—self-similarity—he different constituent parts appear to be like the whole—and scale invariance—fractal structures do not change, independently from the scale in which they are viewed [12, 13]. Therefore, it is the reason why many examples from the natural world could be categorized as

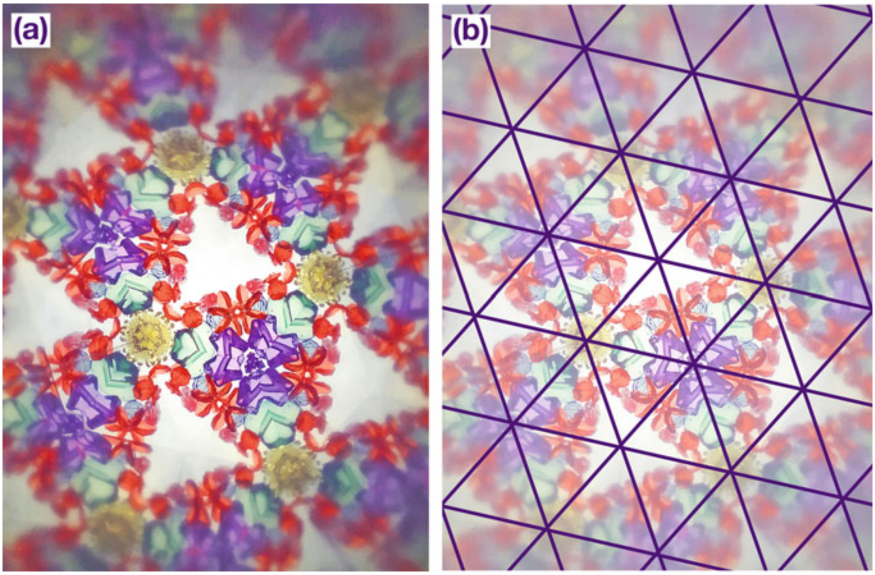


Fig. 1 Difference between "kaleidoscopic image" and "kaleidoscopic structure": **a** Kaleidoscopic image that can be seen through the classic kaleidoscope, in which various types of kaleidoscopic shapes are observed; **b** Kaleidoscopic structure is the organizing structure of the kaleidoscopic image, in most of the cases an isometric grid. (Photo by Vergara, 2019)

kaleidoscopic. However, this geometry does not explain all of the cases proposed within the chapter, such as creating the universe's mirrored representations.

In Sect. "2", examples of microscopic forms like diatoms or snowflakes will be presented, while in Sect. "3", the structure of viruses is studied regarding Buckminster Fuller's architecture influence. In Sect. "4", the morphology and the innovative case of designing new proteins will be introduced concerning the kaleidoscopic condition. Finally, Sect. "5" contains both the kaleidoscopic cosmological models and other kaleidoscopic representations of the universe in astronomical contexts, displaying the kaleidoscopic image as the main motif in this regard.

Please note that the shape of microscopic forms are more similar to the images that can be seen through the classic kaleidoscope, while the vision we might have of a kaleidoscopic universe would be related to the teleidoscopic kaleidoscope view.

2 The Structure of Microscopic Organisms and the Snowflakes

There are many cases in which forms of nature present fractal geometry and, at the same time, show a significant kaleidoscopic nature. About the structure of complex microscopic organisms, also including the radiolarians that Ernst Haeckel illustrated

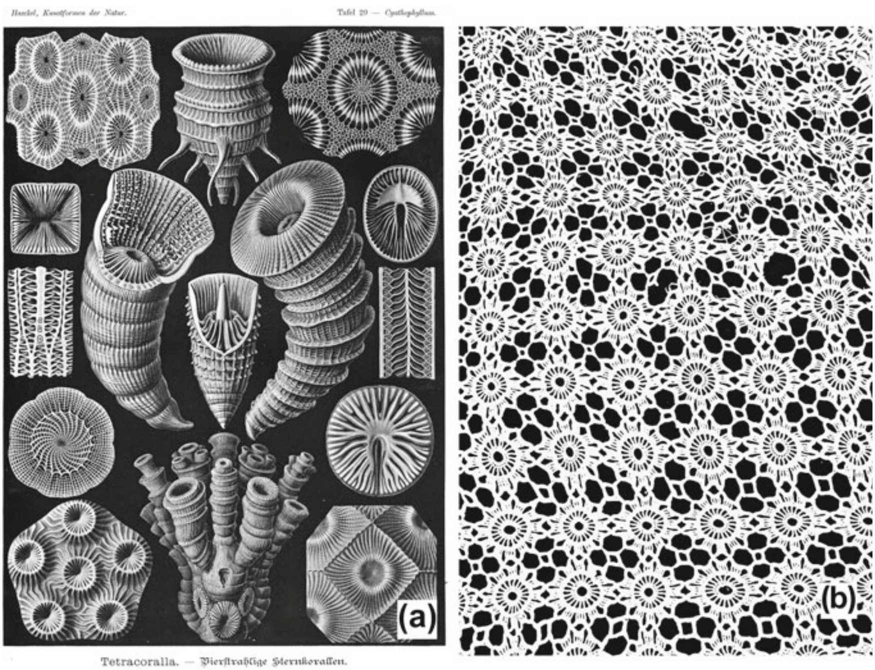


Fig. 2 Visual similarity between the images: **a** Ernst Haeckel, Tetracoralla in *Kunstformen der natur*, 1904 (Adapted from www.biolib.de GNU Free Document License); **b** Crocheted kaleidoscopic pattern. (Photo by Vergara, 2017)

at the end of the nineteenth century, Frank Wilczek specifies that they "*embody the Platonic solids*" referring to their shape. Also, he mentions that the radiolarians are very primitive organisms. For example, the oldest fossils that have been found corresponding to radiolarians that still inhabit our oceans [14]. Ernst Haeckel was the naturalist who created the term Ecology [15], representing in *Kunstformen der Natur*, published in 1904, different organisms that in most of the cases presented radial symmetry. For this reason, many of his engravings are kaleidoscopic and even visually reminiscent of the crocheted kaleidoscopic pattern [6]. This issue can be seen in the comparison of Figs. 2 and 3, displaying the visual similarity between the tetracoralla structure in Haeckel's illustration (Fig. 2a) and the mentioned crocheted pattern (Fig. 2b).

From a biological perspective, the most developed species, such as humans, have evolved to bilateral symmetry. The presence of the radial structure that we can see in kaleidoscopic compositions or the radiolarians mentioned above are usually related to very primitive organisms that have not undergone further evolutionary development [16]. Even though it is considered that the radial symmetry of an organism refers to its primitive condition and, in consequence, is an indicator of being less advanced, this perspective is currently being revised. Recent research indicates that the condition to evolve toward bilateral symmetry depends on the movement of organisms, as well as their relationship with conditions such as gravity [17]. For example, sessile filter-feeding organisms commonly live in a marine environment where gravity is less important, and since they do not require an active displacement like other animals, they did not evolve toward a bilateral geometry model. Considering the radial symmetry of many plants, it is observed that they do not move either and lack a nervous system so that it would be the same case in different conditions. As seen in Fig. 2, most of the organisms illustrated by Haeckel presented radial symmetry, including radiolarians, diatoms, jellyfish, or echinoderms. All of them present a kaleidoscopic morphology and, in the same way, also Haeckel's illustrations do.

In Haeckel's scientific illustration, we commonly find the diatoms, a type of microalgae that usually present a kaleidoscopic structure. If the diatom's shape is not kaleidoscopic, they are visually similar to the small color pieces found inside the classic kaleidoscope tube. The nineteenth century witnessed many kaleidoscopic examples created by scientists from the Victorian era. They were dedicated to the diatom art, almost coinciding in time with the kaleidoscope's invention by David Brewster in 1815. These scientists organized different shapes that presented these microscopic organisms in a microscopic plate to create kaleidoscopic compositions thanks to the diatom arrangement. Klaus Kemp, known as a diatomist, still makes these compositions at present. He was very interested in the diatom arrangement, so he investigated the method to make it. According to Kemp's declaration in the online short documentary on Kemp's work directed by Mathew Killip, *The diatomist*, the Victorian scientists did not leave any clue to discover how to keep diatoms fixed when organizing these compositions [18]. As it is evident in

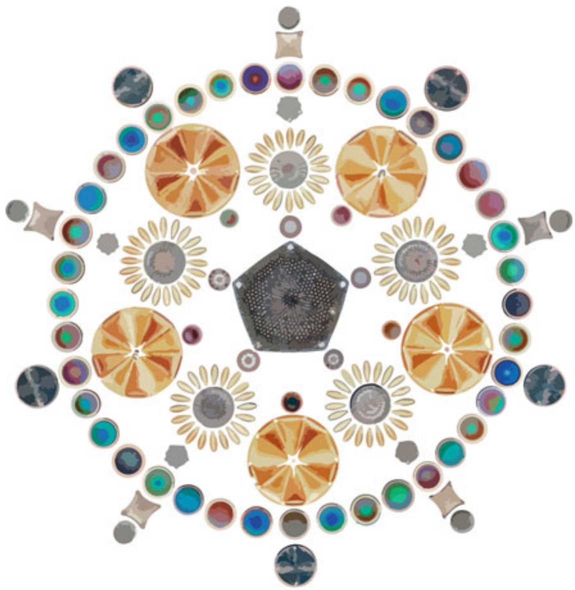


Fig. 3 Illustration based on Johann Diedrich Möller's diatom arrangement displaying the kaleidoscopic organization that the diatoms presented in the nineteenth century. Each item is a diatom. (Made by Vergara, 2020)

the images produced by Kemp or other examples from the nineteenth century, like the high number of images composed by Johann Diedrich Möller[3] (Fig. 3), the arrangement of diatoms into a kaleidoscopic manner is very interesting, considering that the diatoms also present a kaleidoscopic shape on their morphology. Certainly, the visual equivalence between these diatom art images and the vision through the kaleidoscope is very high. Considering that the Victorian scientists started to produce these arrangements after the invention of the kaleidoscope, surely, they received the influence of this device to make their microscopic kaleidoscopic images.

Not only were diatoms arranged kaleidoscopically, but also the snowflakes had been organized in this way by Wilson Alwyn Bentley (Fig. 4a), who made his first snowflake photomicrograph in 1885 [19]. Together with William H. Humphreys, he is the author of a 1931 publication, where a total of 2453 snowflakes are collected. Each snowflake is totally different since the possibility of variations is infinite [20]. He is mostly recognized for capturing different designs, having photographed around five thousand snowflakes throughout his life [19].

Every snowflake within the arranged composition presents a kaleidoscopic shape based on a repetitive fractal pattern similar to the Koch curve due to the snow-crystal radial symmetry with the hexagon as the geometrical basis, as it happens in kaleidoscopic structure. The kaleidoscopic snowflakes arrangement is organized over the isometric grid (Fig. 4b), which was present inside the kaleidoscope (Fig. 1b). At the same time, each one of the snowflakes presents a hexagonal symmetry individually, so the image results to be kaleidoscopic in a double sense.

Like the image constructed by the kaleidoscope, Bentley stated that there is no repetition in the snowflakes' design. That was why he started taking photomicrographs since he thought that the beauty of snowflakes was forever lost without any record [19].

Also, Laure Albin Guillot, born after Bentley, made great artistic contributions in microphotography at the beginning of the twentieth century. As in the case of Bentley, her artistic production had the purpose of crossing borders between art and science, naming their photomicrographs as decorative micrographies instead of microphotography. Furthermore, the results of many of her micrographs are similar to some vortographies took through the vortoscope by Alvin Langdon Coburn. He invented the vortoscope, an optical device that was very similar to the kaleidoscope [9]. What differentiates Albin Guillot from the diatom art compositions already presented or Bentley's microphotography is her intentionality since her purpose was fully artistic and decorative—even though she used elements that traditionally belonged to the scientific field. Among the elements that appear in the

[3] In this case, the figure included that shows the structural disposition of the diatom arrangement made by Möller present a pentagonal organization to continue the structure marked by the central item, not a hexagonal one as it was mentioned in relation to the kaleidoscope. Nonetheless, the kaleidoscopic sense remains present in the composition. Anyway, Möller and other authors have a large production with cases that correspond to hexagonal symmetry as it happens in the most common kaleidoscopic structure.

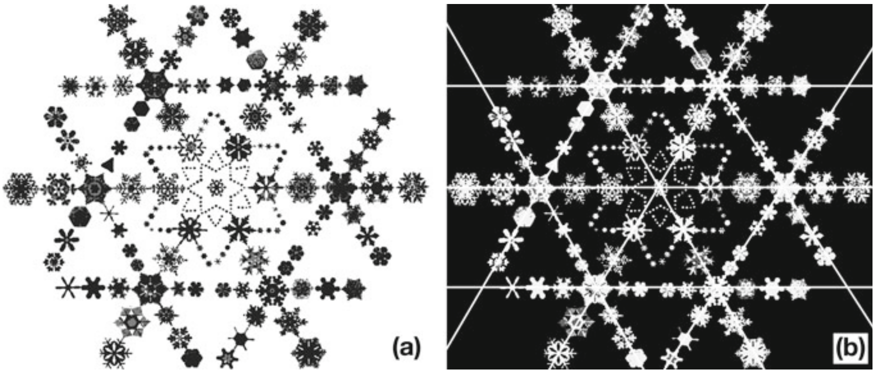


Fig. 4 Illustrations based on a kaleidoscopic composition produced by Wilson Alwyn Bentley: **a** The illustration displays the position of the arranged snowflakes; **b** In the illustration, the kaleidoscopic structure that organizes the composition over an isometric grid is indicated. (Made by Vergara, 2020)

microphotographs made by Albin Guillot, she precisely used the diatoms. Nevertheless, she did not compose the popular kaleidoscopic compositions from nineteenth-century diatom art. Instead, she was interested in showing the kaleidoscopic structure of the diatoms.

In addition to microphotography, Albin Guillot also developed other types of photographic work. About the kaleidoscopic image, her photograph entitled *Les tierces alternées* is very notable. It consists of a kaleidoscopic composition in which multiple repeated hands playing the piano could be seen. One more time, this photograph also remembers the vortographs made by Coburn in the first two decades of the twentieth century [9]. She made it at the end of her career, in 1948, to illustrate *Préludes*, a musical composition by Claude Debussy [21].

3 The Kaleidoscopic Geodesic Dome Designed by Buckminster Fuller and the Morphology of Microscopic Forms Like Viruses or Fullerenes

Buckminster Fuller was the architect who designed and built the geodesic dome in the Black Mountain College (North Carolina) in 1948. It is highly noted that Fuller built the geodesic dome for the first time in Black Mountain College, an idealistic experimental faculty to study arts with a transdisciplinary approach. Figures such as John Dewey, Albert Einstein, or Carl Gustav Jung composed its advisory council, among others [22]. Concerning the procedures used in Black Mountain, Josef Albers—the artist who established a teaching curriculum in 1933—said that *"works of art"* were experiments to gather experience [23]. In Black Mountain, we find kaleidoscopic patterns in artworks made by Ruth Asawa, who studied at Black Mountain, and Anni Albers, the weaving and art teacher. Fuller's geodesic dome presents a kaleidoscopic structure since it is generated by the repetition of equilateral triangles that generate an icosahedron. Regarding this polyhedron, Fuller projected the vertexes to appear as a circumscribed sphere [24].

Regarding Fuller's geodesic dome, its geometric similarity with Islamic art should be noted. Although it does not present any decorative pattern, if a geodesic dome is compared to the Islamic architecture domes, there is a structural correspondence between these two cases. For example, William Harvey's drawings on the structure of the Alhambra's domes are very similar to the geodesic structure. In this regard, Fuller thought that the tetrahedron was the most suitable polyhedron to measured volume instead of the cube: "*he felt that the very nature of space requires the tetrahedron to supplant the cube as the unit of space*" [24]. Certainly, the tetrahedron has the property of covering the entire space through various combinations, closer to the kaleidoscopic structure, since each one of its faces is an equilateral triangle. At the same time, the tetrahedron presents a better geometrical correspondence for the stacking issue in regular polyhedral configurations, for example, regarding the combinations with the octahedron and the icosahedron, than the cube.

Many viruses present icosahedral symmetry concerning their structure, which is also present in Fuller's geodesic dome [24]. It is notable how Fuller's work finally inspired scientists from 1962 when Caspar and Klug published their research on regular viruses' capsids, inspired both in polyhedral figures and the geodesic dome [25]. In their publication, there are images of the geodesic dome and the icosahedral grid that was also the structural pattern in the kaleidoscope, viewed as a plain isometric grid, concretely in the section dedicated to icosahedral viruses [25]. The arranging consisted of icosahedral capsids, which are quite kaleidoscopic concerning their structure (Figs. 5, 6, 7 and 8 in the publication by Caspar and Klug [25]). Afterward, they triangulated the sphere to visualize the shape of viruses' capsids, starting from the isometric grid, designated as "*equilateral-triangular plane net*" within the article, inspired by Fuller's architectural work. They even included a geodesic dome photograph (Fig. 5 in the publication by Caspar and Klug [25]). Recent research in this field showed that Fuller's work still serves as inspiration, considering the previous work by Caspar and Klug, among other authors [26].

Along with the geodesic dome's geometry, we also find an allotropic form of carbon molecule with a truncated icosahedral structure named "Buckminsterfullerene" to honor Buckminster Fuller, which was discovered serendipitously [27]. Fullerenes could work as capsids for other molecules and atoms and were discovered by Kroto et al. in 1985 [28]. The fullerene is formed into a lattice of hexagons and pentagons that seems like a soccer ball [24]. Certainly, these geometrical patterns are particularly kaleidoscopic since the same cell is repeated all along with the "spherical" shape. Carbon appears under different allotropes because the atoms are born together to form different networks [24]. All of these allotropic forms are quite kaleidoscopic as well. In line with the structure of viruses and fullerenes, proteins also present a kaleidoscopic configuration, as discussed in detail in the next section.

These carbon allotropic structures appear in nature, presenting different types of kaleidoscopic grids. Recently, complex fullerenes found in outer space have become the basis for a theory that proposes fullerenes C_{60} found in meteorites to be responsible for carrying extraterrestrial substances that might have started life on the Earth [29]. What could make it possible is fundamentally due to the fullerene configuration as a cage, which would encapsulate atoms or molecules inside. In this sense, it is worth mentioning that Buckminster Fuller thought of possible connections between polyhedral configurations and nature, speculating that "*the icosahedron is immersed within the harmony of the spheres and serves as the origin of life itself*" [24]. If the theory mentioned above were finally proved, Fuller's testimony would acquire an exceptional meaning. In the same way, to conclude this section, Fuller's work was very important not only for architectural and artistic purposes but also for science development concerning viruses' structure and allotropes of carbon, which is very significant regarding transdisciplinary approaches.

4 The Kaleidoscopic Image and the Protein Structure: Designing New Conformations in Protein Folding

Regarding the relation in which kaleidoscopic image and visual arts were involved as interrelated fields of study in the twentieth century, many artists entered the scientific sphere, getting inspiration from discoveries and innovations related to this field. However, it should be noted that artists were preeminent in showing fractal geometry in their artworks many years before fractal geometry was discovered [12]. Throughout the twentieth century, the relationship between science and art has evolved from merely inspiration or scientific photography and illustration to a deeper connection between these areas. Nowadays, the connection between art and science has become progressively more transdisciplinary. For example, this relationship can be found in protein folding, the art of designing new models of proteins with scientific purposes, applying the artistic composition in the kaleidoscopic structure of proteins to create new configurations that did not exist before.

The protein folding and its relation to kaleidoscopic configurations are clear from a visual viewpoint. In 2016, the research on proteins developed by Jacob B. Bale, David Baker, and other authors was published in the journal *Science* [30]. The new designs of proteins, whose application is principally addressed to the field of health and medical issues, can allow the treatment of illnesses or genetic disorders, to separate compounds like gluten from food, etc. For the moment, their functions are simpler than those since the research field of protein folding is almost unexplored considering the possibilities that the design of proteins could offer. Along with this field's development, immeasurable new molecules could be designed like the infinite images that appear within the kaleidoscope. This way, the combinations in protein folding would be related to molecules with unimaginable purposes that do not exist at present. According to Baker, the director of The Institute for Protein Design, usually, the new proteins findings refer to proteins already present in nature and, consequently, with no design involved, which is very unsatisfying. The proteins from nature are a bit limited compared to the "protein universe" that could be newly designed. In other words, he refers to *"all the proteins that could possibly be* *made with varying combinations of amino acids"* [31], which remembers to the infinite condition of the kaleidoscope.

As we see, the design of proteins is moving beyond a transdisciplinary paradigm with a revolutionary approach. The designs are multiple like it happens in the kaleidoscopic device, characteristic for presenting the same structure displayed in the planar representation of an icosahedral surface like an isometric grid. This representation is similar to the mentioned geodesic dome and, at the same time, served as a structure in the illustration for the kaleidoscopic cover in the journal *Science* that was published on July 22 in 2016, Vol. 353, issue 6297. This image, named *Self-assembling protein cages*, was computationally designed by the science illustrator Valerie Altounian [32]. It represents a kaleidoscope that aims to illustrate the protein folding issue in relation to the articles published by Bale et al. [30] and Robert F. Service [33], respectively, named as "Accurate design of megadalton-scale two-component icosahedral protein complexes" and "Rules of the game." Visually, the image contains *"120-subunit icosahedral protein nanoparticles (yellow, dimers; blue: trimers; green, pentamers)"* for creating a kaleidoscopic array [32], based on protein models provided by Bale [34]. Therefore, the colored pieces that are *inside the kaleidoscope* that is displayed on the cover, at the same time, create a pattern that visually refers to the proteins. At the same time, the kaleidoscopic image on the cover is distinguishing proteins by colors, aiming to represent the nanoparticles that form the different protein cages. Compared to the photomicrographs from the article by Bale et al. [30], where most of the proteins present an icosahedral axis, proteins and both kaleidoscopic image and structure displayed in the cover are very similar from a visual viewpoint.

According to the *Science* statement on the cover, the combination of pieces within *"these shapes result in a kaleidoscopic array of self-assembling protein complexes, each of which rivals the size of a small viral capsid"* [34]. These proteins, assembled in a kaleidoscopic way, *"could form the basis for a new generation of biomolecular machines"* with customizable applications [34]. In consequence, the design of kaleidoscopic as the changing shapes within the kaleidoscope opens the transdisciplinary paradigm called *the protein folding revolution*. There is an interesting explicative video where Altounian's kaleidoscopic illustration is animated [35]. This way, the vision of a kaleidoscopic image is also defined as its property of being a continuously changing shape.

5 The Presence of Kaleidoscopic Image as the Main Motif in Cosmological Models and Other Representations of the Universe

The astronomical community is interested in the kaleidoscopic approach, both regarding cosmology and the universe's representation. In fields such as Astrophysics or Cosmology, it has been found how there is a kaleidoscopic representation system to visualize the universe in this way.

5.1 Kaleidoscopic Cosmological Models: The Geometry of Finite Topologies that are Visually Infinite

Similar to the truncated icosahedron shape of the fullerene or other examples seen above, within this section, we principally refer to another Platonic solid: the dodecahedron. This regular polyhedron has regular pentagons as faces instead of equilateral triangles, as in the case of the icosahedron, but it could be considered as kaleidoscopic for being fully composed by the conjunction of pentagons. In 2003, based on the background microwave radiation, an article was published indicating that the universe could take the form of the dodecahedral space, based on Poincaré's geometry in contrast with the standard cosmological model of an infinite flat universe. This cosmological model, proposed by Luminet et al. [36], presents a multi-connected condition, which is very close to the kaleidoscopic nature.

This theory proposes a visually infinite universe within spatial boundaries. Then, it would be like being inside an infinite space, as it would happen if we could enter as observers in between the mirrors of a kaleidoscopic finite space. In fact, the first author of the study, the astrophysicist Jean-Pierre Luminet, refers to the kaleidoscopic nature of the space to introduce the dodecahedral universe issue in one of his publications, in the section entitled as *The Hall of Mirrors*. Concretely, he introduces the multi-connected topology of Poincaré's dodecahedral space and other related models, referring to a cubic one to introduce this issue. To do this, he uses the metaphor of a room entirely covered by mirrors *"on all six surfaces"* where *"a kaleidoscopic effect will be produced"* [37]. Within this type of multi-connected topology, *"any object in space may possess several copies of itself"* [37], as it happens in the kaleidoscopic image. According to this theory, there would be a possibility of being into a kaleidoscopic universe. Luminet explains that cosmic space could immerse us in that same illusion without the presence of walls or mirrors. In this case, the kaleidoscopic nature would be generated by the *"multiplication of the light ray trajectories following the folds of a wraparound universe"* [37]. Thus, he refers that we could be *"in a physical space which is closed, small and multiple-connected"* and still being able to see a larger observed space *"like a part of a tessellation made of repetitions"* [37], falling into an illusion even though our view would come from the mirrored repetition of a "fundamental cell" where the observer is located [37]. This description is extremely interesting dealing with the kaleidoscope, a very small, closed, and physical space that visually generates an infinite virtual space. While the image that is seen through the kaleidoscope could be categorized as fictional, the observer is involved like it were a real view, which is effectively produced by the repetition of an element that comes from reality [7, 11]. The reference to the tessellation is very notable as well since the kaleidoscope displays an image that is very similar to the patterns from decorative arts, as was introduced in Sect. "1".

The reason for presenting this theory is that it is quite significant for visual studies and, obviously, about the kaleidoscopic nature. Regarding the existence of a kaleidoscopic dodecahedral universe, it is worth noting that Plato referred to the dodecahedron as the form that contains the whole universe, while the other four Platonic solids correspond to the four elements: water, air, earth, and fire [38].

In 1900, the architect Eugène Hénard designed *The Hall of Illusions* [39], which was the first surrounding kaleidoscopic space [9]. The same year, the astronomer and physicist Karl Schwarzchild, suggested the existence of a kaleidoscopic cosmology, characteristic for referring to a finite space with boundaries that at the same time generated the illusion of infinity [40]. In this way, Schwarzchild refers to a kaleidoscopic cubic space made of identical repetitions of the Milky Way galaxy, tilling an infinite three-dimensional space that remembers to Escher's creations like *Cubic Space Division (Cubic Space Filling)*: *"One could imagine that as a result of enormously extended astronomical experience, the entire universe consists of countless identical copies of our Milky Way, that the infinite space can be partitioned into cubes each containing an identical copy of our Milky Way. Would we really cling on to the assumption of infinitely many identical repetitions of the same world? We would be much happier with the view that these repetitions are illusory, that in reality space has peculiar connection properties so that if we leave any one cube through a side, then we immediately reenter it through the opposite side"* [40].

Escher's images are intended to suggest the coexistence of simultaneous worlds [41], so it is not strange to find a correspondence between his work and Schwarzschild's statement. Together with Escher, both Hénard's *Hall of Illusions* and the dodecahedral universe theory are analogous to the condition expressed by Schwarzschild. Indeed, Luminet has found inspiration concerning this quote for the dodecahedral universe theory [36], as shown in an older publication by Luminet and Roukema [42]. In the end, all of these references reflect visually infinite spaces, although they are finite and present boundaries, which correspond with the kaleidoscopic nature. As Schwarzchild stated within his proposal, to cross through one of the boundaries would lead the subject to reenter into the same space from the opposite side. In this line, it is essential to refer to Jeff Weeks' work, the second author of the dodecahedral cosmological model [36], whose dedication is geometric topology and cosmology.

Although at present it is not known which one is the cosmological model that our universe presents—Weeks mentions the fact that the universe could be finite or infinite—some proposals for visualizing these cosmological models have been considered [43]. Interestingly, Weeks refers to the multiple copies of the Milky Way, which in a certain way could work as a reminiscence of Schwarzchild's suggestion. Weeks goes further and refers to the enormous complexity of this matter. When the universe is observed, the speed of light factor should be considered. On the opposite case of the identical images that could be seen through the kaleidoscope, within a kaleidoscopic universe, the observer will be watching not the present but the past in any direction [43].

In Schwarzchild's line, Weeks introduces that one of the possibilities is to be part of a finite universe with boundaries that at the same time could be visually infinite [43]. This is that something could be infinite and, at the same time, present a boundary as it happens in the kaleidoscope. Once again, this formulation coincides with the hypothesis presented on the dodecahedral universe [36]. To visualize this type of spaces, among other applications, Weeks designed the flight simulator *Curved Spaces* [44, 45], whose software—which is available for computers and mobile phones—allows the subject to interact within a space where different geometric topologies of multi-connected universes can be explored like a multiverse. These universes are visually infinite, although they present boundaries, so the repetition of the same cell generates the kaleidoscopic image. The options for visualizing the universe in this way correspond to flat, hyperbolic or spherical topologies. It also allows seeing a kaleidoscopic view with no walls, having the Earth, a galaxy, or a gyroscope as the main repetitive motif, or to close the walls of the room to get an understanding of the space morphology. The person can interact within the walls, so the structure could be seen with geometric lines that divide the different kaleidoscopic cells.

Curved spaces software allows the visualization of many geometric models with the kaleidoscopic universe as the central motif. For example, the model of the dodecahedral universe and other topologies can be visualized as well. Although the geometrical models are different, in general, all of them share the same characteristics of being multi-connected and visually infinite cosmologies. Together with the mentioned issue, there is the possibility of being inside mirrored and non-mirrored universes. About the kaleidoscopic image, the visualization of mirrored geometries is more suitable for being kaleidoscopic than the non-mirrored ones. In this way, the subjects could be participants of their juxtaposition of ubiquities within the self [7]. It is allowed thanks to a spaceship view that indicates the position where the observer is situated, which corresponds to multiple views of the kaleidoscopic universe where the subject is repeated.

In addition to the aforementioned *Curved Spaces*, there are two applications designed by Weeks: *Kaleidopaint* [46, 47] and *Kaleidotile* [48, 49]. As it is indicated in the title, *Kaleidopaint* allows the subject to produce kaleidoscopic designs similar to those that could be found in a decorative pattern. Thus, when something is painted, the pattern becomes simultaneously painted within the rest of the composition kaleidoscopically. *Kaleidotile* makes it possible to visualize kaleidoscopic designs on polyhedral shapes and Euclidean or hyperbolic planes, as well as the transition between different polyhedra—for example, to transform a dodecahedron into an icosahedron.

In every case, there are kaleidoscopic motifs on the faces of the polyhedron or the plane. In consequence, the transformation of these shapes also means to visualize the kaleidoscopic image continuous metamorphosis. Without a doubt, it is striking that Weeks designed his software both to visualize kaleidoscopic cosmological models and to design kaleidoscopic images in different applications. In this sense, a remarkable correspondence is detected in a double association: the kaleidoscopic visualization of the universe and the design of kaleidoscopic images with different purposes—mathematical, geometric, artistic, etc.

5.2 Kaleidoscopic Representations of the Universe in Astronomical Contexts: *Zero Gravity—Multi Mirror Projection Room* (ESA, 2010) and *GTC Kaleidoscope* [50]

As presented in the previous section, kaleidoscopic cosmological models have been proposed. Consequently, it is not strange to find resources where the universe was represented kaleidoscopically, specifically in astronomical dissemination contexts. These spaces, whose main constituents are mirrors, generate multiple repeated images characterized by their kaleidoscopic nature. Furthermore, these characteristic mirrored spaces offer the observer the possibility of seeing, or even being totally immersed, within a kaleidoscopic space. Subsequently, this experience may produce the feeling of floating inside a multiverse, where the subjects are together with their own copies. For this reason, the study of the universe's kaleidoscopic representations that have been found in Spain, Europe, has been considered very relevant.

At the City of Arts and Sciences in Valencia was where one of these kaleidoscopic immersive spaces had been found. It is a video installation that forms part of the permanent exhibition *Zero Gravity* at the *Science Museum* and is organized by the European Space Agency (ESA). The immersive kaleidoscope is named *Multi Mirror Projection Room*, having been exhibited in the museum since 2010. In a certain way, this kaleidoscopic, virtual and immersive space shares similarity with the *Hall of Illusions* designed by Hénard in 1900. However, instead of infinitely reproduce the appearance of the room that became an ornamented palace, there is not a repeated architectural motif or decoration, but an image commonly related to the universe view (Fig. 5). Therefore, it is more similar to the perspective introduced by Luminet regarding the kaleidoscopic hall, as well as to the Weeks' software for the visualization of an infinitely multi-connected universe. The *Zero*

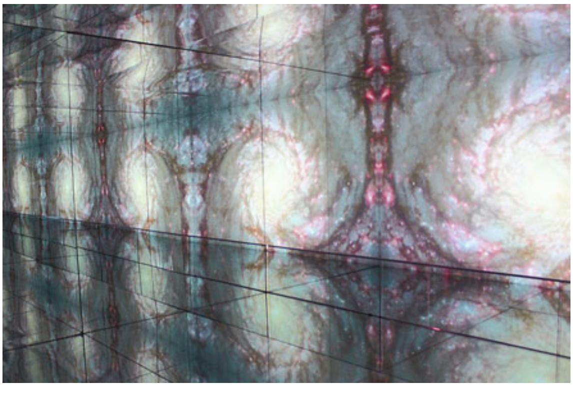


Fig. 5 Multi mirror projection room (*Zero Gravity*, ESA, 2010): a visible pattern that shows an infinite kaleidoscopic universe. The kaleidoscopic image motif corresponds to the Whirpool galaxy (NASA, ESA, STScI, AURA). (Photo by Vergara, 2019)

Gravity mirrored cube video installation works perfectly to experience Luminet and Weeks' proposals.

This immersive kaleidoscope is entirely composed of mirrors: all the walls, the floor, and the ceiling are completely covered. Images of the universe are projected over one of the walls, specifically over a 237" screen with 16:9 format that at the same time presents an optical reflection mirror where aerial photographies of planet Earth, astronomical objects, astronauts, or space stations are observed, among other motifs.[4] Therefore, a kaleidoscopic pattern is generated by the projected motif mirrored repetitions and the reflections of the people that are inside the room that works as the central cell. As a consequence of being inside a mirrored room of considerable dimensions,[5] the subject participates in a multisensorial space where being infinitely repeated, floating within a multi-connected universe (Fig. 6). Given this circumstance, the observers are immersed inside a kaleidoscopic universe where everybody is seeing themselves in every direction.

Concerning the kaleidoscopic image (Fig. 7a), the room structure is also very important. The mirrors that cover all the room are rectangular and right-angled at the corners. For this reason, the view is visually like the cubic geometries in *Curved Spaces*. For example, if the observer looks to the corner, the view is quite similar to the visualization of a three-dimensional multi-connected torus (Fig. 7b).

A prototype of the immersive kaleidoscope has been studied as well. Since it is a video installation model, it is named *Zero Gravity—Multi Mirror Projection Room Model*. It is located at the European Center for Space Astronomy (ESAC) in Villanueva de la Cañada, Madrid, although in this case, it is not an immersive kaleidoscope but a small cube covered by mirrors. The observer could look into the inner cube to see a visually infinite universe. Within the cube, two videos produced by ESA were projected: on the one hand, the same video that was projected in the kaleidoscopic immersive room, *Zero gravity* (2010), and *Our Changing Planet* (2007), which showed images that mostly belonged to the ESA archives and is not projected in the mirrored room video installation.[6]

[4] In the video, most of the images' source of origin are the ESA and NASA images archives, although the video installation also presents images from the following organisations: CNES, ARIANESPACE, GSFC, METI, ERSDAC, JAROS, Silicon Worlds, MPS, OSRIS Team, UDP, LAM, IAA, RSSD, INTA, UPM, IDA, DLR, FU Berlin (G. Neukum), JPL, Space Science Institute, University of Arizona, SOHO-EIT, STScI, AURA, PACS & SPIRE Consortium, HOBYS Key Programme Consortia.

[5] Some technical information on the *Zero Gravity—Multi Mirror Projection Room* video installation: The floor area in the projection room is, approximately, 43 m^2. The dimensions of the mirrored room are the size of 79.57 m^2 (13.26 $\times$ 6 m) and 3 m as height. From the outside, the room has the appearance of a diagonally leaning cube, so its height ranges from 3.74 to 5.36 m. Video length: approximately 7 min.

[6] The sources of the images that appear in the video *Our changing planet* (ESA, 2007): ESA, Europe's metereological missions (Meteosat, Meteosat Second Generation, Meteosat Third Generation, MetOp), ESA's environmental missions (ERS-1, ERS-2, Envisat, Proba-1), Earth Explorers (GOCE, SMOS, CryoSat-2, Adm-Aeolus, Swarm, EarthCARE), GMES dedicated missions (Sontinel-1, Sontinel-2, Sontinel-3, Sontinel-4, Sontinel-5), Envisat, Technical University of Denmark, NASA.

Fig. 6 Multi mirror projection room (*Zero Gravity*, ESA, 2010). The mirrored room where the subject's multiple reflections could be observed. The kaleidoscopic image motif corresponds to the planet Earth (Envisat, ESA). (Photo by Vergara, 2019)

The prototype reproduces the mirrored room on a minor scale, where the mirrors infinitely repeat the images projected over one of the cube faces. Both the immersive kaleidoscope that is suitable for the human scale as well as its prototype model generates a kaleidoscopic experience by reflecting an original motif—the one projected over one of the cube faces—along with the copies of the people that are inside the room, in the case of the immersive kaleidoscope (Fig. 8). Likewise, the kaleidoscopic image's fractal condition is notable in relation to these spaces since it means to visualize fractal configurations from nature—for example, galaxies or fractal elements from aerial views of planet Earth—in a kaleidoscopic way (Figs. 5 and 7a). In other words, the subject inside the mirrored room participates in a "double fractal" view, which is produced both by the kaleidoscopic structure and due to the fractal condition that the projected images present.

Compared to the kaleidoscopic room, few variations are when the *Zero Gravity* video is played in the prototype. Inside the model, the video length is slower than in the immersive room. Also, the music is different in each case. In the immersive video installation, the *Gymnopédies*, composed by Erik Satie, is played, while in the ESAC prototype, there is relaxing and repetitive ambient music.

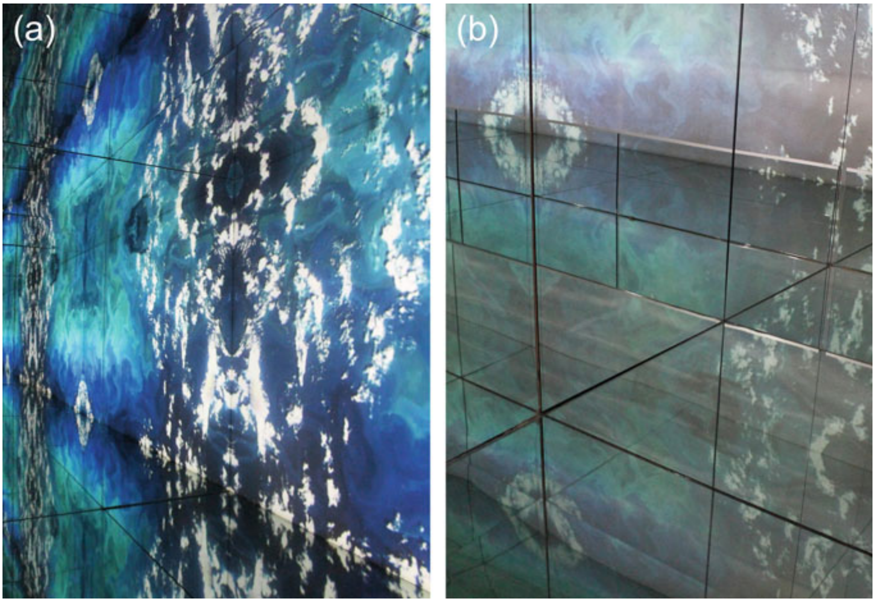


Fig. 7 Multi mirror projection room (*Zero Gravity*, ESA, 2010). The kaleidoscopic image motif corresponds to the plankton proliferation in the Barents Sea (ESA): **a** Pattern that shows an infinite kaleidoscopic image; **b** Structure formed by the mirrors, whose view in the corner is quite similar to a three-dimensional multi-connected cubic torus. (Photo by Vergara, 2019)

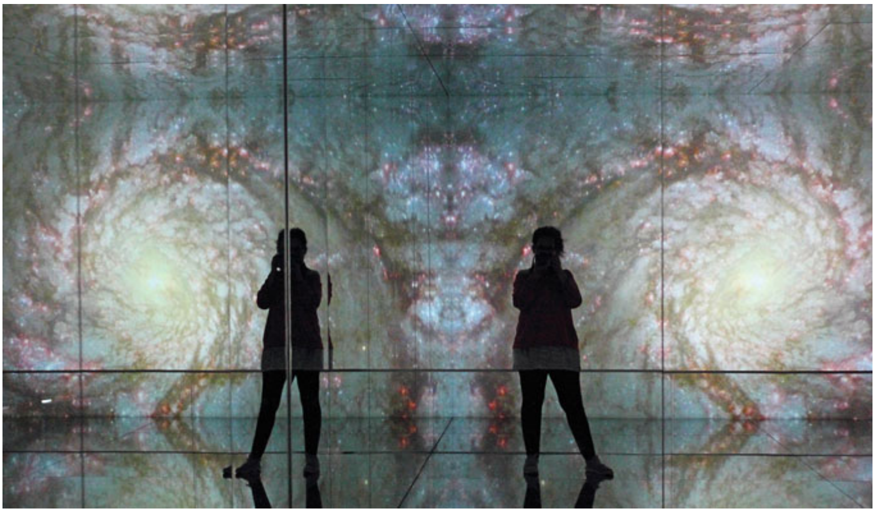


Fig. 8 Multi mirror projection room (Zero Gravity, ESA, 2010). Mirrored cube where the subject's multiple reflections could be observed. The kaleidoscopic image motif corresponds to the Whirpool galaxy (NASA, ESA, STScI, AURA). (Photo by Vergara, 2019)

Apart from these issues, the main difference between the room and the model is related to the subject's multisensory experience. The kaleidoscopic room incorporates the subject as part of the video installation, while the kaleidoscopic image that could be seen inside the prototype is observable, but the reflections do not show the observer as part of the kaleidoscopic space.

With a similar configuration of the *Zero Gravity—Multi Mirror Projection Room* (ESA, 2010), another case has been found. It is the *GTC Kaleidoscope* video installation at the Canary Islands Astrophysics Institute (IAC). Specifically, the *GTC Kaleidoscope* video installation was produced in 2015 by the IAC for the exhibition *Lights of the Universe: 30 Years of the Canary Telescopes* in Tenerife and La Palma. The title refers to the Gran Telescopio Canarias (GTC), the Observatory Telescope, located at Roque de los Muchachos, La Palma. *GTC Kaleidoscope* was an immersive space that, once more, generated the view of a kaleidoscopic universe where the projected video was repeated infinitely. In this case, we find the direct reference to the term "kaleidoscope," which replaces the word "telescope." Within the description of the *GTC Kaleidoscope* video installation, it is mentioned that the visualization of this type of space in cosmological models refers to the view of a kaleidoscopic universe [50]. For this reason, it worked as an immersive video installation whose principal aim was to experience the kaleidoscopic infinite universe, similar to the *Multi Mirror Projection Room* in the *Zero Gravity* exhibition (ESA, 2010), which is still displayed at present.

6 Conclusion

The kaleidoscopic image manifests a multiscale nature since it is present from microscopic forms to the mirrored multi-connected representations of the universe. Surprisingly, it follows the condition enunciated in an ancient proverb: *As above, so below. As below, so above*. This is not only because of the similar morphologies and topologies that have been categorized as "kaleidoscopic" in this chapter. But also, there is a correspondence between the micro- and the macro-cosmos views when the kaleidoscopic image is involved. At the same time, it should be noted that the shape of these forms is based on repetition, producing a mirrored pattern.

Overall, the visualization of the kaleidoscopic image and its structure corresponds both to the scientific and the artistic field, being an important topic for the visual studies area. In conclusion, it is a visual configuration whose geometry is extensively present in fractal examples from nature, as well as structurally displayed in proposals of multi-connected cosmologies. Definitely, this image is suitable for being investigated from many perspectives and disciplines with different purposes, so it should be considered for developing transdisciplinary approaches. The kaleidoscopic paradigm as an intersection between art and science would bring and open new perspectives, discoveries, and possibilities for both areas of knowledge, as can be seen in transdisciplinary research like protein design.

Core Messages

- In nature, there are similar morphologies and topologies whose shapes result from the repetition, producing a mirrored pattern categorized as "kaleidoscopic".
- The kaleidoscopic image is a visual configuration present from microscopic forms to the mirrored multi-connected representations of the universe.
- The nature of the kaleidoscopic image is multiscale, showing the correspondence between the micro- and the macro-cosmos views.
- The visualization and research on the kaleidoscopic image and its structure correspond both to scientific and artistic fields.
- The study of the kaleidoscopic image as an intersectional element opens a transdisciplinary paradigm that would bring new perspectives and discoveries.

Acknowledgements The author would like to thank the expertise collaboration of the astrophysicist Dr. José Antonio Caballero, who is a researcher at the Center for Astrobiology (CAB) (National Institute of Aerospace Technology [INTA], The Spanish National Research Council [CSIC]) at the European Space Astronomy Centre (ESAC), Spain, and Dr. Juan J. Vergara, Professor of Marine Ecology at University of Cádiz (UCA), Spain. The author also would like to thank the supervision of Dr. José María Parreño and Dr. Daniel Silvo on the issues presented within this chapter during the development of her Ph.D. Thesis (Complutense University of Madrid 2019) and expresses her gratitude for the attention she received from Rosa Martí at the Sciences Museum (The City of Arts and Sciences, Valencia).

References

1. Woolf V (1985) Moments of being. Harcourt Brace and Company, New York
2. Denes A (2020) Agnes denes. Manifesto 1970. Available via agnes denes' website. http://www.agnesdenesstudio.com/works15.html. Accessed 10 July 2020
3. Drożdż S, Oświęcimka P, Kulig A, Kwapień J, Bazarnik K, Grabska-Gradzińska I, Rybicki J, Stanuszek M (2016) Quantifying origin and character of long-range correlations in narrative texts. Inf Sci 331:32–44. https://doi.org/10.1016/j.ins.2015.10.023
4. Vergara MN (2019) The kaleidoscopic experience within every-day-life: on authorship, intertextual narrative, stream of consciousness and multifractality. Popular inquiry 2:12–22. Available via Aaltodoc website. http://urn.fi/URN:NBN:fi:aalto-202001021003. Accessed 10 July 2020
5. Müsch I, Seba A (2015) Gabinete de curiosidades naturales. Taschen, Barcelona
6. Vergara MN (2017) La perspectiva caleidoscópica y la asociación visual como metodología de investigación artística. In: Facultad de Bellas Artes (ed) 1ª Jornada PhDay Complutense Bellas Artes. Universidad Complutense, Madrid, pp 110–116. Available via Eprints Complutense. https://eprints.ucm.es/50393/. Accessed 10 July 2020
7. Vergara MN (2020) The kaleidoscopic interference and the double-slit project: borges' thought, quantum behaviour and multiverse theory. Artnodes 25:1–12. https://doi.org/10.7238/a.v0i25.3317

8. Crary J (1990) Techniques of the observer: on vision and modernity in the nineteenth century. MIT Press, Cambridge
9. Vergara MN (2018) La imagen caleidoscópica como primera forma de abstracción en los inicios del arte moderno. ANIAV 3:80–95. https://doi.org/10.4995/aniav.2018.10068
10. Brewster D (1819) A teatrise on the kaleidoscope. Archibald Constable and Company, Edinburgh
11. Vergara MN (2019) Visualizando lo invisible: La imagen caleidoscópica y los procedimientos asociados a la visualidad líquida, múltiple y conectiva. ASRI 17:52–69. Available via ASRI. https://www.eumed.net/rev/asri/17/visualidad-liquida.html. Accessed 10 July 2020
12. Mandelbrot B (2009) La geometría fractal de la naturaleza. Tusquets, Barcelona
13. Garrido H, García Ruíz JM (2009) Armonía fractal de Doñana y las Marismas. Lunwerg, Barcelona
14. Wilczek F (2016) El mundo como obra de arte. Drakontos, Barcelona
15. Deléage JP (1993) Historia de la ecología. Icaria, Barcelona
16. Martindale MQ, Henry JQ (1998) The development of radial and biradial symmetry: the evolution of bilaterality. Am Zool 38(4):672–684. https://doi.org/10.1093/icb/38.4.672
17. Holló G (2015) A new paradigm for animal symmetry. Interface Focus 5(6):1–10. https://doi.org/10.1098/rsfs.2015.0032
18. Killip M (2014) The diatomist. Available via Vimeo. https://vimeo.com/90160649. Accessed 10 July 2020
19. Blanchard DC, Bentley WA (1998) The snowflake man: a biography of Wilson A. Bentley. The McDonalds and Woodward Publishing Company, Blacksburg
20. Gombrich EH (1999) El sentido del orden: estudio sobre la psicología de las artes decorativas. Debate, Madrid
21. Tachon F (2020) Albin Guillot laure. Available via France's culture ministry website. https://www.culture.gouv.fr/Sites-thematiques/Musees/Nos-musees/Valorisation-des-collections/Les-femmes-artistes-sortent-de-leur-reserve/Icones/Albin-Guillot-Laure. Accessed 10 July 2020
22. Miguel V (2015) Looking back at the impact of black mountain college. Artsy, October 26. Available via Artsy. https://www.artsy.net/article/the-art-genome-project-looking-back-at-the-impact-of-black-mountain-college. Accessed 10 July 2020
23. Asawa R (2020) Black mountain college. Available via Ruth Asawa's website. https://www.ruthasawa.com/life/black-mountain-college/. Accessed 10 July 2020
24. Kappraff J (2001) Connections: the geometric bridge between art and science. World Scientific, Singapore
25. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor Symp Quant Biol, New York 27:1–24. https://doi.org/10.1101/SQB.1962.027.001.005
26. Cepelewicz J (2017) The illuminating geometry of viruses. Quanta magazine, July 19. Available via Quanta magazine. https://www.quantamagazine.org/the-illuminating-geometry-of-viruses-20170719. Accessed 10 July 2020
27. Kroto H (1993) The birth of C_{60}: buckminsterfullerene. In: Kuzmany H, Fink J, Mehring M, Roth S (eds) Electronic properties of fullerenes. Springer series in solid-state sciences, vol 117. Springer, Berlin, Heidelberg, pp 1–7. https://doi.org/10.1007/978-3-642-85049-3_1
28. Kroto HW, Heath JR, O'Brien SC, Curl RF, Smalley RE (1985) C_{60}: buckminsterfullerene. Nature 318:162–163. https://doi.org/10.1038/318162a0
29. Instituto de Astrofísica de Canarias (2013) Las moléculas más complejas del universo. Available via IAC website. http://www.iac.es/es/divulgacion/noticias/las-moleculas-mas-complejas-del-universo. Accessed 10 July 2020
30. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, Baker D (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. Science 353(6297):389–394. https://doi.org/10.1126/science.aaf8818
31. Zimmer C (2017) Scientists are designing artisanal proteins for your body. The New York times, December 26. Available via The New York Times. https://www.nytimes.com/2017/12/26/science/protein-design-david-baker.html. Accessed 10 July 2020

32. Altounian V (2016) Self-assembling protein cages. Available via Valerie Altounian's website. https://www.altounianillustration.com/portfolio/self-assembling-protein-cages. Accessed 10 July 2020
33. Service RF (2016) Rules of the game. Science 353(6297):338–341. https://doi.org/10.1126/science.353.6297.338
34. Science (2016) On the cover. Science 353(6297):325
35. Science Magazine (2016) The protein folding revolution. Available via the Youtube account of Science Magazine. https://youtu.be/cAJQbSLlonI. Accessed 10 July 2020
36. Luminet J-P, Weeks JR, Riazuelo A, Lehoucq R, Uzan J-P (2003) Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background. Nature 425:593–595. https://doi.org/10.1038/nature01944
37. Luminet J-P (2008) The shape and topology of the universe. In: Proceedings of conference tessellations: the world a jigsaw, Leyden (Netherlands), march 2006. Cornell University Library, New York, pp 1–21. Available via arXiv preprint. https://arxiv.org/abs/0802.2236. Accessed 10 July 2020
38. Giomini C (2003) Timaeus's insight on the shape of the universe. Nature 425:899. https://doi.org/10.1038/425899c
39. Sazatornil L (2012) De Diebitsch a Hénard: El 'Estilo Alhambra' y la industrialización del orientalismo. In: Calatrava J, Zucconi G (eds) Orientalismos. Arte y Arquitectura entre Granada y Venecia. Abada, Madrid, pp 53–72
40. Schwarzschild K (1998) On the permissible curvature of space. Class Quantum Gravity 15(9): 2539–2544
41. Ernst B, Escher MC (2007) El espejo mágico de MC Escher. Taschen, Madrid
42. Luminet J-P, Roukema BF (1999) Topology of the universe: theory and observation. In: Lachièze-Rey M (ed) Theoretical and observational cosmology. NATO science series (Series C: mathematical and physical sciences), vol 541. Springer, Dordrecht, pp 117–156. https://doi.org/10.1007/978-94-011-4455-1_2
43. Weeks J (2013) Jeff Weeks: "Shape of Space"—Aalto university MathArt colloquium. Available via the Youtube account of Aalto university. https://youtu.be/j3BlLo1QfmU. Accessed 10 July 2020
44. Weeks J (2020) Curved spaces. Available via geometry games website. http://www.geometrygames.org/CurvedSpaces/index.html.en. Accessed 10 July 2020
45. Weeks J (2020) Curved spaces. Available via app store. https://apps.apple.com/us/app/curved-spaces/id1417178571. Accessed 10 July 2020
46. Weeks J (2020) Kaleidopaint. Available via geometry games website. http://geometrygames.org/KaleidoPaint/index.html. Accessed 10 July 2020
47. Weeks J (2020) Kaleidopaint. Available via app store. https://apps.apple.com/us/app/kaleidopaint/id518275953. Accessed 10 July 2020
48. Weeks J (2020) Kaleidotile. Available via geometry games website. http://www.geometrygames.org/KaleidoTile/index.html.en. Accessed 10 July 2020
49. Weeks J (2020) Kaleidotile. Available via app store. https://apps.apple.com/us/app/kaleidotile/id1374849164. Accessed 10 July 2020
50. Instituto de Astrofísica de Canarias (2015) Luces del universo. Available via IAC website. http://www.iac.es/es/divulgacion/noticias/luces-del-universo. Accessed 10 July 2020

Mari Nieves Vergara (Málaga 1990) holds a Ph.D. in Fine Arts with the Extraordinary Ph.D. Award (Thesis defended in 2018-2019 at the Fine Arts School, Complutense University of Madrid, Spain) and International Mention, whose thesis title was *"The Juxtaposition of Ubiquities in Kaleidoscopic Image."* In autumn semester 2016, she was granted an Erasmus + fellowship for a postgraduate Ph.D. stay in the Department of Art of the School of Arts, Design and Architecture at Aalto University in Helsinki, Finland, where she collaborated as a researcher and visiting teacher. She has a Degree in Fine Arts (University of Málaga 2012) and finished her studies at Malmö University, Sweden, in an Erasmus annual exchange. She has completed a Master's Degree in Artistic Research and Creation and a Master's Degree in Teaching in Visual Art Education (Complutense University of Madrid). As a visual artist, she participated in exhibitions in Spain and had been awarded for her artwork. Mari Nieves has published her scholarly work in books and academic journals such as ANIAV, ASRI, Popular Inquiry, and Artnodes.

10 Disruptions and Competitive Strategies in the Automotive Industry

Dušan Marković and Veljko Mijušković

> "*The difficulty lies, not in the new ideas, but in escaping from old ones*".
>
> John Maynard Keynes

Summary

The new business environment, created as a result of accelerated globalization at the beginning of the twenty-first century, is characterized by frequent economic and technological disruptions. These disruptions unfroze competitive relationships in many global industries and resulted in their restructuring. Great Recession, technology innovations, changing customers' preferences, stricter environmental regulation, and the ongoing crisis of coronavirus disease (COVID-19) have forced companies in traditional mature industries, such as the automotive industry, to adapt competitive strategies to the new reality. Previously established sources of competitive advantage are no more sufficient to survive in the automotive industry. Therefore, competitors in the industry have to manage and balance the processes of exploration and exploitation. Auto-companies have invested the lion's share of their revenues in R&D of electric, connected, and autonomous vehicles, but uncertainties about R&D performances are still very high. To face mounting the R&D expenditure,

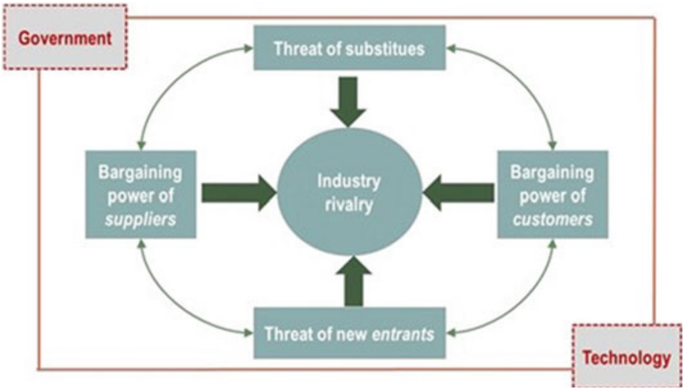
D. Marković (✉) · V. Mijušković
Faculty of Economics, Kamenicka 6, Belgrade, Serbia
e-mail: dusan.markovic@ekof.bg.ac.rs

V. Mijušković
e-mail: veljko.mijuskovic@ekof.bg.ac.rs

D. Marković
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Belgrade, Serbia

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_10

decreasing demand, and fierce competition in the industry, they need to restructure the supply chain and increase production efficiency. To achieve a balance between cost-efficiency and differentiation, many auto-companies establish strategic alliances with direct competitors and digital suppliers and acquire incumbents and innovative startups.



Extended five forces model (Porter in Harv Bus Rev 86:79–83, 2008).

The code of this chapter is 01000001 01110101 01110100 01101111 01101101 01101111 01110100 01101001 01110110 01100101 00100000 01101001 01101110 01100100 01110101 01110011 01110100 01110010 01111001.

1 Introduction

The last decade of the twentieth century and the beginning of the twenty-first century have witnessed the business environment's rising globalization. Technological, economic, and social disruptions in this period inspired a large number of studies about the future of global business. Economic interdependence between developed and emerging markets, technological innovations in telecommunications and transportations encouraged some researchers to conclude that the world's future is complete globalization [1]. Other researchers think that Brexit, US protectionism trade policy, the USA–China trade and technological war, and current COVID-19 crises are signs of slowing globalization [2], which will result in a new business environment with a larger influence of national governments on business activities. However, neither of these predictions is completely accurate because today's business reality is semi-globalization, where foreign factors significantly impact national economies. Still, the influence of domestic factors is larger and more important [3].

Globalization has different effects in developed and emerging markets. Asian middle class's real income has grown significantly due to globalization, while the western middle class's real income growth has been negligible [4, 5]. High growth rates and rising consumers' real income in emerging markets have motivated many western multinational companies (MNCs) to establish production plants and transfer modern technology solutions. However, they face huge political and business risks in these markets [6]. Foreign direct investments boost the economic development of emerging markets, attracting more and more foreign investors. Moreover, foreign factors' influence is not the same in every industry, and some regions inside a country are more available to foreign companies than others. This situation is common in emerging markets. In emerging markets, governments limit and strictly control foreign investments in strategically important industries and stimulate foreign investments in some underdeveloped regions trying to boost benefits and nullify globalization costs [7].

Many governments perceive the automotive industry as global and strategically important because the industry, directly and indirectly, creates millions of jobs, significantly contributes to tax revenues, and stimulates export and foreign direct investments. Additionally, the industry is capital-intensive and technology-intensive, and there are significant spillover benefits on local companies [8]. It is why many governments apply a broad stimulus to attract auto-companies to invest in their countries.

The structure of the auto-industry was stable, and there were only gradual changes governing competitive relationships. However, ever since the Great Recession, the industry has been facing disruptive processes. These processes are shaking the industry, resulting in a new competitive reality and the emergence of technological pioneers and competitors from emerging markets.

This paper analyzes how economic and technological disruptions have changed the automotive industry structure and how auto-companies have adapted their competitive strategies to this new reality. We use "The five forces model" introduced by Michal Porter, the most powerful and widely used tool, to analyze the industry's structure and competitive pressure. The model consists of five forces: (i) rivalry inside an industry, (ii) suppliers, (iii) buyers, (iv) substitutes, and (v) new entrants, whose strength defines industry profitability [9]. When coming under criticism, the researchers point out that the model is static and does not include any influences of technology and government actions on an industry's structure [10]. Porter partially accepted the critiques and included two additional forces in the model, but he argued that the new forces' influence on industry profitability is only indirect. Technology and government actions affect the five forces' strength, not the industry profitability directly [11].

The following part of this chapter studies the Great Recession's impact on the automotive industry structure and how auto-companies restructure their business models to adapt to the changing global demand, fierce competition, the pressure that comes from the emerging market late followers, and government actions. We then analyze how disruptive innovations in digitalization, connectivity and electrification, stringent ecological regulations, and growing global geopolitical turmoil have changed the industry structure and competitive strategies. Finally, we provide some new data and analyze what consequences the current COVID-19 global crisis may have in the industry.

2 The Influence of the Great Recession on Automotive Industry Structure

The Great Recession began in 2007 as a financial crisis in the USA, but later on, it spilled over worldwide. US expansive monetary policy and low-interest rate policy at the beginning of the twenty-first century resulted in growing liquidity in financial markets, which encouraged banks to issue mortgage loans to subprime borrowers. This created a bubble in the real estate market and caused a sharp rise in financial leverage in the financial sector. When the bubble burst, many banks faced financial problems, which caused severe liquidity contraction and sparked crises of confidence [12]. Stricter borrowing criteria, rising interest rates, and slowing economic activity resulted in a sharp rise in unemployment and plunging demand. The Recession was the deepest economic crises and the severest that many developed countries had ever faced. The transformation of some mature and asset-heavy industries, such as the automotive industry, was triggered by the Recession.

The auto-industry was mature, and the industry structure was stable before the Recession. Competitive relations, customer preferences, and technology were changing gradually, enabling MNCs from developed markets to build high entry barriers and reap benefits from pioneering [13]. However, the industry experienced considerable changes during the Recession. The changes are due to a sharp decline in demand, unpredictable customer preferences, government actions, and new competitors' emergence in the premium segment.

Few MNCs from developed countries controlled the lion's share of the market and set industry trends (Fig. 1), while companies from emerging markets produced cheap and low-quality products for domestic price-sensitive customers. The industry was capital and technology-intensive, so cost-efficiency and economies of scale were the primary competitive advantage sources. Many auto-companies applied a global strategy for achieving cost-efficiency [14]. This strategy produces minimal adaptation to local customers' preferences. Production and R&D functions had been centralized in domestic markets to achieve economies of scale before incumbents started to apply regional strategies. The regional approach means the establishment of production plants in emerging and transition markets. The plants were established in low-cost production locations to serve demand on Triads[1] markets and balance production and logistic efficiency. Low value-added activities in the middle of the supply chain have been outsourced to emerging and transition markets [15], while high value-added downstream and upstream activities were centralized in developed markets. The regional strategy has enabled auto-companies to adapt slightly marketing strategies to the local business environment and sustain cost-efficiency [16]. High capital intensity and the importance of economy of scale were efficient entry barriers that deterred late followers from entering the market. However, when demand plunged due to the Great Recession, the incumbents faced overcapacity and low productivity. Consequently, General Motors, and Chrysler filed for bankruptcy protection, while Ford divested from

[1] Western Europe, North America and Japan.

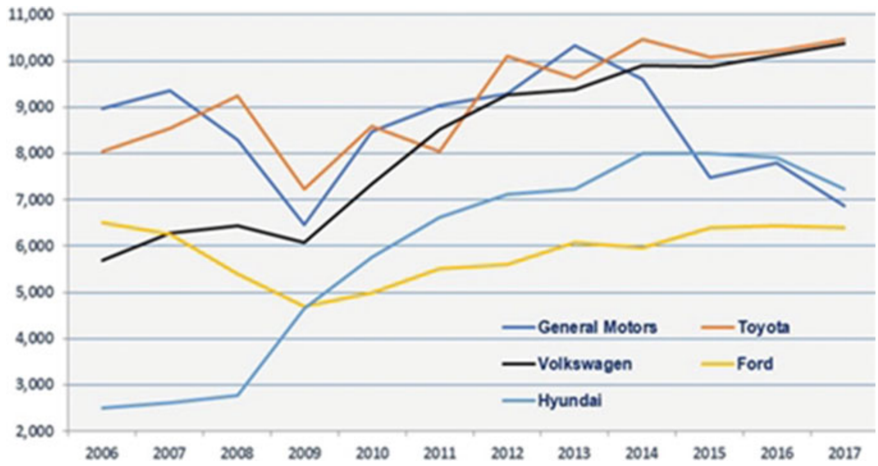


Fig. 1 Main competitors in the automotive industry—in thousands of vehicles [19]. It shows the volume of production for main auto-companies in 2006–2017. (Prepared with data from OICA http://www.oica.net/category/sales-statistics/)

Volvo and Jaguar Land Rover and focused on the domestic market to survive. Other auto-companies also experienced severe financial problems but managed to survive the Recession.

Emerging markets' late followers used the turmoil in the industry and acquired troubled automakers from developed companies [17]. The late followers used acquisitions to obtain intangible resources like cutting-edge technology, know-how, marketing and management competencies, and valuable brands necessary to fill the competitive gap [18]. By combining their low-cost production capabilities with the acquired resources, late followers could survive the Recession and make competition fiercer. On the other hand, acquisitions realized during the Recession by companies from developed markets had other motives. For example, Fiat acquired bankrupted Chrysler in 2010 to penetrate the North American market and to acquire the competencies necessary for serving the segment of larger models. Additionally, Fiat used the acquisition to achieve economies of scale, improve production efficiency, and adjust production capacity to the new reality.

Traditional sources of competitive advantage in the auto-industry, namely agility, flexibility, and efficiency [20], strongly depend on a company's ability to organize and manage the global supply chain. The complexity of design and supply chain management, capital and technology intensity of the industry, brand loyalty, and the importance of balancing regulatory and customer demand are entry barriers that prevent new entrants' emergence during the Recession [21].

Auto-companies have a leading role in developing product architecture, designing modular platforms, and system integration, while other activities are outsourced to suppliers [22]. Supply chain performances are affected by the competencies to build business networks with key suppliers to foster trust and knowledge sharing inside the network [23]. Perez and Sanchez [24] argued that auto-companies with strong relationships with suppliers could achieve higher productivity and quality than competitors reporting weaker alliances with suppliers.

During the last decade of the twentieth century and the beginning of the twenty-first century, consolidation among the suppliers resulted in the emergence of "mega suppliers." Auto-companies outsource more value-added activities to these first-tier suppliers but retain tight control and key system integration activities, even if the suppliers' innovative capabilities have not been fully exploited. Because of that auto-companies retained the lion's share of the value created inside the supply chain [25].

The Recession affected suppliers' bargaining power. Magna, a first-tier supplier, tried to acquire Opel, a General Motors subsidiary, when the US government nationalized General Motors. The aim was to acquire valuable brands, distribution channels, and system administration capabilities and penetrate the industry with higher margins. Magna's key customers were against the acquisition. Volkswagen threatened to break up Magna's cooperation if it acquired Opel because Magna would be its direct competitor. Because of strong pressure, Magna postponed the acquisition and focused on its core competencies. On the other hand, auto-companies considered the Recession as an opportunity to absorb more value created inside the supply chain and lower suppliers' pressure. The Fiat Chrysler acquisition mentioned above aimed to restructure the supply chain and make it more efficient. After the acquisition, Fiat and Chrysler joined together their supply, reconsidered relationships with suppliers, made long-term arrangements with Asian top-quality suppliers, decreased the number of modular platforms, and worked on developing new platforms. After the acquisition, Fiat achieved an economy of scale in production and supply, which resulted in a growing sale and improved its financial performance.

Demand in the industry strongly depends on economic growth, while customer preferences differ on the regional level. For example, the US market is the main market for pick-up vehicles, while the European market is the main market for small city models. Before the crisis, the global market reached its historical peak, with 71.5 million sold vehicles. The main regional markets, i.e., US, EU, Chinese, and Japanese markets, made up almost 70% of global demand [26]. Due to the long-lasting stagnation of real income before the Recession [5], the plunge in real income during the Recession, and growing uncertainty, customers in developed markets postponed purchasing long-lasting products, including vehicles [27]. During the Recession, the US auto-market shrunks by almost 40%, and the EU market shrunks by 16%. US market dropped sharply during the crises, but the recovery was strong, while the EU market drop was modest due to government scrap subsidies. The subsidies were used in many European countries to boost demand and recover the domestic auto-industry [28]. On the other hand, during the crises, the Chinese market surged from 7.3 million vehicles in 2006 to 19.4 million vehicles in 2012, becoming the largest national market. The Chinese market's strong growth attracted many auto-companies from developed countries, boosting investment and technology transfers, resulting in fierce competition and benefits for customers.

Declining real income and growing uncertainty resulted in a growing demand for "value for money" products, not only in emerging but in developed markets, too [29]. Even consumers who did not experience a decrease in real income change their preferences to cheaper and simpler products [27] or energy-efficient alternatives. US auto-companies, known as producers of large energy-inefficient vehicles, were not flexible and could not adjust their product portfolio to the new reality, so they offered significant discounts to the customers [30]. This approach was not successful because it could not sustain demand, and prices went down, resulting in financial troubles and bankruptcy. Changing demand was a business opportunity for companies with capabilities to produce smaller, energy-efficient, and "good enough" vehicles. Hyundai–Kia exploited these capabilities to expand into the growing "value for money" segment in developed markets, where one can perceive its brand as low price and low quality. To change brand perception, the company invested in production technology to sustain cost-efficiency, improve quality, extend the warranty, and modernize product design [31]. Due to the growing quality, positive customer experience, and acceptable prices, Hyundai–Kia global sales jumped from 4 million vehicles in 2008 to 7.7 million vehicles in 2014.

The pressure of substitutes was very low in the auto-industry before the Recession. Key alternatives to owning a car were public transportation and cycling. During the Recession, governments, especially in developed countries, banned diesel vehicles from city centers or imposed additional taxes [32]. They tried to demotivate customers from owning a car, encouraging them to use alternative ways of transportation. Governments financed projects which promote cycling as a healthier and environmental friendly alternative too. These projects targeted tiny but growing customer segments to avoid automotive transport limitations, such as rising costs of owning a car, shortage of parking places, air pollution, and road traffic injury [33] (Table 1).

We thoroughly analyzed the Great Recession's impact on the automotive industry structure and concluded that the Recession changed the business environment and faced auto-companies with growing threats. Demand plunged, competition became fiercer, suppliers tried to absorb more value from the supply chain, and customers changed their preferences. All of that led to confusion in the industry. Reaction to the changing structure was consolidation on the global level. The main drivers of cross-border acquisitions in technology-intensive industries are speed to the market, obtainment of new technology, obtainment of complementary technology solutions, acquisition of local marketing and management knowledge, improved production efficiency, etc. [34]. The cross-border acquisition enabled acquirers to reshape business models or improve the efficiency of the existing business model [35].

Acquirers from developed markets used acquisitions to improve the efficiency of the existing models. The main drivers of these acquisitions were achieving economies of scale in production, supply, and R&D, achieving economies of scope, removing excess capacity, and sharing modular platforms. On the other hand, emerging markets multinationals conduct acquisitions to reshape their business models. Their acquisitions of targets from developed markets aim to obtain strategic

Table 1 Structure of automotive industry before and during the great recession

	Before the great recession	During the great recession
Industry rivalry	Couple of/several large MNCs controlled a lion share of the market and set industry trends	Few large competitors filed for bankruptcy protection Late followers from emerging markets used acquisitions to jump into the premium segment Producer of smaller and energy efficient cars penetrated into developed markets
Bargaining power of suppliers	Suppliers were integrated in networks lead by auto companies Consolidation in suppliers' industry	Suppliers tried to exhibit forward vertical integration Suppliers have grabbed more value from value created in a supply chain
Threat of new entrants	Entry barriers were very high preventing emergence of new competitors	Supply chain complexity, technology and capital intensity deter new entrants
Bargaining power of customers	Demand was stable Customers' preferences were different on regional level	Demand plunged in developed markets, while demand in Chinese market soared Significant growth of value for money segment in developed markets
Threat of substitutes	Very low	Small, but there is a growing segment of customers who change preferences toward cycling
Technology	Gradually changing	Gradually changing
Government	Provide stability of business environment	Nationalization and bailout of US auto companies Huge scrape subsidies in Europe Promotion of alternatives to owning a car

It shows how the five forces, technology, and government activities affected the automotive industry structure before and during the Great Recession

intangible resources, such as technology, brands, business models, talents, and innovative skills. The acquired resources come together with low-cost resources and efficient operating processes in domestic markets [36]. The acquirers treat targets more like a strategic partner, pursuing the companies' structural separation, and creating synergy through coordination of their business activities, which results in preserving and upgrading the strategic resources [37]. Previous research found out that emerging markets multinationals' integration is more successful than the approach used by acquirers from developed markets [38].

3 Contemporary Disruptions in Automotive Industry

After the Global Recession, the structure of the auto-industry underwent a significant change. It looked like that the new structure would be stable in the next decade, and auto-companies would have enough time to adapt business models to the new reality and to recover from the Recession. However, the new cycle of disruptions faced financially exhausted auto-companies with new threats and an unpredictable environment.

Even if they are not in direct contact with the industry, disruptive technology innovations could trigger a plunge in the industry's profit and sometimes a termination in the short run [39]. Asset heavy and technology-intensive industries, such as the automotive industry, felt protected from radical and fast changes in the industries' structure. In contemporary business, companies in these industries face a prolonged stealthy weakening of their core business, which results in the slow shrinking of their financial performance without signs of recovery [40]. Technological challenges (ACES: autonomous, connected, electric, and shared/services mobility) and stringent emissions standards changed the impact of five forces in the auto-industry, decreasing the industry's financial performance and encouraging auto-companies to transform core operations [41, 42] and penetrate new business and industries.

Economic expansion in China and stable recovery in developed countries after the Great Recession caused changes in demand in automotive markets. Chinese market grew by more than 55% in 2011–2017 due to the growing middle class (Fig. 2). Slowing GDP growth, economic, and political clashes with the USA resulted in the market's squeeze in 2018–2019. In this period, demand for cheap, low-quality cars plunged, while demand for "value for money" and premium products rose significantly. Demand in developed markets reached pre-crisis levels by the end of 2019 (Fig. 2), while demand structure changed significantly. SUV and cross–over segment have revamped in the US market and have grown significantly in the European market, while demand for smaller, energy-efficient, and value for money segment have dropped.

Digitalization and electrification have driven changes in customer preferences. The goal of electric cars' development is to establish a balance between demand for mobility and environmental protection, which has been significantly harmed by auto-companies' operations. This relatively small market segment has been growing rapidly, but some auto-companies have neglected it. In 2019, almost 2.3 million units of "green" vehicles were sold and made up 2.5% of the global auto market [43]. Technology innovations have not been sufficient for this segment's development although the segment growth is mostly due to environmental regulation. National governments use different policies and tools to boost "green" vehicle demand. These measures include public procurement, financial incentives to buy the vehicle and/or cut the cost of its usage (e.g., free parking), and a large scale of regulatory restrictions [44]. Market factors have contributed slightly to the development of this segment [45]. Governments' financial incentives are insufficient to

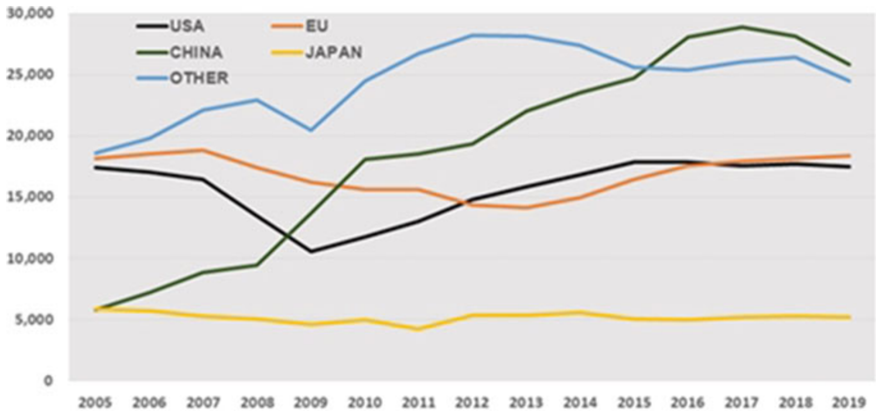


Fig. 2 Global automotive sales—in a million vehicles [26]. It shows the sales in the main automotive markets in 2005–2019. (Prepared with data from OICA http://www.oica.net/category/sales-statistics/)

surge the demand because customers are worried about the limited range of "green vehicles," long charging time, and low infrastructure availability [46]. The future of the segment depends on technological breakthroughs, which will result in significant cost reduction and improved performance of batteries and higher availability and improved performance of public fast chargers [44].

Digitalization enables auto-companies to improve the performance of their products and offer additional services in the markets. Due to digitalization, modern vehicles can communicate with other vehicles and infrastructure, which results in improved safety, more efficient driving, and lower pollution [47]. Government regulation, especially in developed markets, requires auto-companies to provide connectivity for their models because of the wider society's benefits [48]. On the other hand, customers expect modern vehicles to keep them always connected, increase safety, and provide entertainment services during the ride [49]. These additional services are no longer premium, so the customers' willingness to pay extra money for them is declining [48, 50].

Auto-companies have to find a solution to meet growing customer expectations and dealing with stringer regulation. To survive a new unpredictable environment, auto-companies have to become ambidexter firms. Ambidexter companies can efficiently exploit the existing competencies, and at the same time, flexibly explore new competencies, trying to balance cost-efficiency and product differentiation [51]. Ambidexterity results in superior business performance and has a positive impact on firm survival [52]. Exploration and exploitation projects compete for scarce and limited resources, so management has to balance them properly.

To create new competencies, auto-companies are increasing R&D investments in "green technology" and digitalization projects. In 2019, auto-companies and auto-parts producers invested EUR 128 billion, a comparable amount with pharmaceutical and biotechnological industries [53]. The largest auto-companies and premium segment competitors make the lion's share of R&D expenditure in the industry.

The example of Swedish auto-company SAAB shows that building a competitive advantage only on technology innovations without paying attention to business efficiency results in competitive disadvantage and bankruptcy [54]. Companies combine the existing technology with the new one and create hybrid products to deal with the opposite business aims, thus introducing an interim step between proven old and unpredictable emerging technological solutions [55]. In the auto-industry, a good example of this approach is the development of hybrid cars, which combine an electric propulsion system with internal-combustion engines (ICE).

The economy of scope and economy of scale in R&D and production are very important for improving the auto-industry's business efficiency. To achieve the economy of scale and economy of scope, auto-companies decided to expand into market segments where they traditionally had not competed. Competitors from the mass segment introduced independent premium brands, like DS launched by PSA, or Genesis introduced by Hyundai. On the other hand, premium auto-companies expanded into the mass segment, offering smaller and cheaper models, like Mercedes A and B class. Production flexibility has a positive impact on industry cost-efficiency. Flexibility is related to lean thinking and a company's competencies to produce more than one model in one production plant and its ability to transfer a model from one plant to another. Auto-companies invest in cutting-edge production platforms that work in association with the production of a dozen models and the development of digital solutions for production and supply chain. These innovations lead to improved production efficiency and flexibility, which are sources of competitive advantage in the automotive industry.

Technological disruptions, changing customer preferences, and more stringent regulations caused changes in the industry rivalry. Like General Motors, some old leaders have not recovered from the Great Recession and have not managed to restructure their business model, which resulted in declining market power. Toyota and VW have emerged as leaders in the industry, which compete successfully in premium and mass segments and developed and emerging markets simultaneously. In the meantime, followers have been looking for merger opportunities to prepare them for an innovation cycle. In the last couple of years, Nissan acquired Mitsubishi; PSA acquired Opel; PSA merged with FCA, and Renault and Nissan are considering a merger proposal. Consolidation in the automotive industry will result in an economy of scale, a growing risk appetite, and increasing investments in new technologies and new business models.

The automotive industry is an asset-heavy and mature industry characterized by high sunk costs [40], the importance of continuity and resistance to change [56], and the strong impact of brand heritage on consumers' attitudes [57]. In the past, these characteristics were the barriers that kept the newcomers out of the market. However, technological disruptions opened the way for newcomers to enter the market. The first new entrant was the US Company Tesla, which focused on the electric and automotive cars segment. The company uses a micro-entry approach, focusing on the small, growing, and highly profitable segment, which has been neglected by large incumbents [58]. A few years ago, large incumbents considered the segment of electric vehicles too small and unprofitable to invest significantly in electric batteries' development. However, changes in environmental regulations and the potential benefits of corporate social responsibility motivated the main incumbents to enter this segment. Tesla had to craft a unique competitive strategy to beat the well-established incumbents in this segment because traditional auto-companies have complementary resources, such as competencies in engine design, development and production know-how, well-known brands, which can be transferred and additionally exploited in the segment [59]. To prevent traditional companies from exploiting complementary resources, Tesla took a competitive advantage of its innovative corporate culture, along with attempts to leverage suppliers' knowledge to compensate for their limited manufacturing capabilities [60]. Tesla has been trying to expand the segment and build an ecosystem by opening up its patent of charging infrastructure to competitors [22]. New entrants who only target specific and profitable segments and later penetrate larger segments are beginning to emerge [41].

Radical innovations in digital communications, Internet technologies, and electronics caused a restructuring of the automotive industry's supply chain [22], making the supply chain open to incumbents with digital capabilities and expertise [61]. Suppliers with digital expertise are increasing their share in created value in the supply chain. Estimates say that the share of software in the predicted vehicle value rises from 10 to 60% in the near future [49], attracting newcomers from technologically intensive industries. This trend makes digital suppliers' bargaining power stronger and additionally decreases profitability in the industry. The profitability of the industry can be additionally eroded if some of the first-tier digital suppliers decide to acquire an auto-company. In this situation, the suppliers will combine their own technology capabilities with the acquired brand, market linkages, and capabilities to design and organize the production of vehicles [62]. Such acquisitions are not unprecedented in business practices because Microsoft and Google, as software companies, entered the smartphone, i.e., the hardware industry by acquiring Nokia and Motorola, respectively. These acquisitions were not successful, but both companies learned a lot about managing producers of hardware, which enabled them to make future acquisitions successful. To prevent this scenario, auto-companies should firmly integrate the suppliers in the production network, build motivation, trust and commitment [63], and collaborate with suppliers to develop new technologies and capabilities [64]. This approach makes the supply chain more efficient and results in growing interdependence between auto-companies and suppliers, making it harder for suppliers to enter the supply chain's downstream activities.

Technological advances, especially the development of autonomous vehicles, will open the way for real substitutes for owning a car. Digital innovations enable the development of different business models and mobility services, such as ride-hailing, car-sharing, and peer-to-peer, giving more flexibility to consumers. Demographic factors and lifestyle significantly influence consumers' attitudes toward the services, while the vehicle fleet negligibly influences customers' decisions [65]. Many startups and well-established companies have entered the emerging industry, expecting to benefit from pioneering. Still, the consumer acceptance rate in modern mobility services is very low, which results in low asset utilization and the unprofitability of the service providers [48]. Many local (Didi in China or Ola in India) and global providers (Uber) have not given up and are increasing investment in new business models making the competition fierce. Auto-companies consider the development of shared mobility as a potentially significant threat, so some enter this field through partnerships or establish their own service providers. The shortening of the innovation cycle and faster diffusion of the innovations in the next decade will shift the auto-industry's revenue pool toward sharing services [41].

Incumbents in the automotive industry use two different strategies to deal with technological disruptions. The first, largest group comprises companies that compete in the mass and value for money segments (FCA, PSA, Nissan, SAIC etc.). These companies are late followers, waiting for standardization of technology in the industry and investing less than 5% of sales in R&D [53]. Meanwhile, companies are increasing the utilization and exploitation of the existing assets and competencies and improving their business efficiency and profitability. Due to the unpredictable business environment, the late followers limit investment in new technologies, conduct smaller R&D projects, terminate unprofitable brands and market segments, and save cash when the industry's situation becomes stable. These companies are looking for merger opportunities in the industry, enabling them to achieve an economy of scale in production and share risks and R&D expenditure. The late followers expect slow diffusion of the market innovations, which will allow them to exploit complementary resources, invest in standardized technologies, and even acquire some financially exhausted pioneer or fast followers (Table 2).

The other incumbent group comprises leaders in the industry (Toyota and VW) and premium auto-companies (Daimler and BMW). These fast followers invest heavily in digitalization, connectivity, electric, and autonomous vehicles, thus preventing innovative outsiders from penetrating the industry and first-tier digital suppliers from grabbing more value created in the supply chain. For example, VW revealed its plan to invest 44 billion EUR in the e-mobility, autonomous vehicles, and digitalization of products and plants. High investments are not sufficient to fill the competitive gap, so they still lag behind innovative suppliers and agile entrants. Incumbents did not recognize the potentially devastating effects of the technological disruptions on the industry's profitability, so now, they are trying to catch up with the innovative competitors by forming strategic alliances with other fast followers [66]. The main aims of the strategic alliances are to make a significant breakthrough in the development of autonomous, connected and electric vehicles, set the industry standards, share R&D expenditures, and decrease the bargaining power of digital suppliers. The incumbents now have the same business aims, but they have to remember that they are direct competitors, so they have to develop a mechanism to protect their own source of competitive advantage and prevent partners from misusing access to sensitive capabilities [67]. Strategic cooperation with direct

Table 2 The structure of automotive industry after the great recession

Industry rivalry	Incumbents have to balance exploration and exploitation to survive Incumbents have entered a market segment in which they did not operate previously Rising consolidation in the industry
Bargaining power of suppliers	Growing importance of digital suppliers The first-tier digital supplier may acquire troubled auto companies and enter automotive market
Threat of new entrants	Supply chain complexity, technology and capital intensity deter new entrants. Tesla has found way to bypass the barriers and penetrate into a small but growing market segment
Bargaining power of customers	Global demand has been driven by Chinese market Growth of "value for money" segment in emerging markets, and SUV and cross-over segments in developed markets Demand for "green" vehicles is growing Customers expect connectivity and digital services in the car, but they are not ready to pay more for these features, because they are not considered as premium
Threat of substitutes	Rising, but unprofitable industry of modern mobility services. Some auto companies entered the industry
Technology	Radical innovation in the area of connectivity, digitalization and electrification of vehicles
Government	Stringent environmental regulation and incentives to buy "green" cars

It shows how the five forces, technology, and government activities affected the structure of the automotive industry after the Great Recession

competitors is very risky, stressful, and needs specific management capabilities, frequently resulting in low performance [68], so many fast followers react to technology disruptions through acquisitions of the emerging entrants in the industry [61]. The emerging entrants provide cutting-edge technologies and capabilities to design and develop innovative solutions, while incumbents provide funds, valuable brands and market linkages, and share the knowledge on how to launch and commercialize innovative products in the market. The main obstacle to an acquisition's success is the efficient integration of fluid organizational structure and vibrant corporate culture of emerging companies into traditional auto-companies' formal organizational structure. Companies that can balance the emerging companies' innovativeness with their own financial strengths, mechanical excellence, and marketing and management knowledge will create synergy and sustain a competitive advantage.

In 2020, global business, including the automotive industry, faces a looming economic crisis caused by the fast diffusion of COVID-19. The disease emerged in China, but swift COVID-19 has spread worldwide, resulting in a plunge in global economic activities. The IMF predicts a decrease of advanced economies by 8%, emerging markets and developing economies by 3%, and a slow recovery in 2021 [69]. The plunging economy has resulted in growing unemployment, especially in developed economies. For example, the USA lost 19.6 million jobs in February–May 2020, so the employment growth from the last eight years was nullified [70].

As we have already highlighted in this chapter, the auto-industry is susceptible to economic turmoil, and this crisis is not the exception to the rule. Growing unemployment, decreasing real income, and increasing uncertainties have resulted in diving demand for vehicles. EU demand for passenger cars dropped by 41.5%, and demand for commercial vehicles dropped by 36.7% in the first five months of 2020 [71]. Additionally, governments have imposed many travel and trade restrictions to prevent disease spread, threatening to break up global supply chains and enhance the crises' negative impact.

Financially exhausted, struggling with technological disruptions, auto-companies have to find a way to cope with the emerging economic crisis. Restructuring plans of some auto-companies aim to cut operating costs, terminate excess capacity, focus on core markets and marketing segments, and reshape the supply chains. Nissan announced the termination of European operations; Renault cut off about 20,000 jobs globally, including 4,600 jobs in France, and Ford redesigned its business strategy, focusing on selling the most profitable models like SUV, pick-ups, and light commercial vehicles. Severe operating cash flow problems caused by decreasing sales and large R&D expenditures, which cannot be cut easily, have forced some auto-companies to seek financial support from national governments. Renault was the first such company that secured the government's loan of 5 billion EUR in June 2020 to stay afloat during the pandemic. These restructuring plans also highlight the importance of reshaping supply chain strategies and sourcing more from local markets. It is especially the case for large markets because it helps auto-companies manage the rising governments' pressure for local content [72] and cope with the growing trade tariffs and restrictions, threatening to create bottlenecks supply chains globally.

The COVID-19 economic crisis is at its beginning, and it is tough to predict its future development and its full impact on the auto-industry. The only certain thing is that flexible companies that can transform their business models, following the technological, market, and economic disruptions, will survive this turmoil.

4 Conclusion

The auto-industry was growing and developing in a relatively stable and predictable business environment for several decades. The industry's profitability was high, and entry barriers were formidable, deterring outsiders from penetrating the market. The industry was mature and asset-heavy and seemed safe from radical changes, but ever since the Great Recession, auto-companies have been facing constant disruptions, which radically transformed industry structure. Global economic crisis, radical and frequent technological innovation, stringent pollution regulation, and fast-changing and unpredictable customer preferences have decreased industry profitability.

The Great Recession caused a significant negative impact on the automotive industry. During the Recession, demand plunged, especially in the US and European markets, and even large financial incentives given by governments could not stop it. Additionally, customers changed their preferences toward smaller and energy-efficient vehicles. Unagile and inflexible auto-companies, which did not have capabilities to restructure business operations and adapt product portfolio to the new reality, faced severe financial problems and even bankruptcy. In response to all this, auto-companies tended to acquire troubled incumbents and restructure their business operations. The restructuring plans consisted of the termination of excess capacity in developed markets, investment in modular platforms, which offer the production of different models and brands, and investment in emerging markets with strong growth during the Recession. Some late followers from emerging markets used the industry's turmoil to acquire debt-burdened companies from developed markets and thus penetrate developed markets and premium segments.

Few years after the Recession, when it seemed that a better time for the industry was coming, auto-companies faced new disruptions. Technological disruptions such as autonomous cars, digitalization, connectivity, and electrification of vehicles have opened the door to the automotive industry for outsiders who operate in the field of consumer electronics and digital platforms [61]. The outsiders have been attacking small but profitable segments, such as the electric and autonomous vehicles segment, while digital suppliers have been trying to grab the lion's share of the value created in the supply chains.

Faced with the disruptions and technological pioneers' competitive pressure, the incumbents have decided to be late followers or fast followers. Both groups try to balance exploration and exploitation in the long run but focus more on one or another type of activity depending on the technological cycle phase. Fast followers are currently giving more importance to exploration, trying to fill the competitive gap. To obtain new competencies necessary to survive disruptions, they have invested a significant amount of money in R&D projects in the digitalization and electrification area, formed strategic alliances between themselves, and acquired innovative startups and digital suppliers. Fast followers' success depends on their ability to efficiently combine traditional and obtained digital resources, coordinate digital transformation with competitive strategy, and bring innovative corporate culture in their organization [73].

Late followers exploit the current resources in the center of their competitive strategy, but they do not neglect the exploration of new technologies and capabilities. They carefully invest scarce resources in smaller R&D projects and upgrade complementary resources waiting for technological standardization. They aim to have a strong financial structure when the technology becomes standardized and then significantly invest in technology and marketing, and aggressively attack pioneers and fast followers. To implement this strategy, it is necessary to achieve an economy of scale and economy of scope, so fast followers conduct horizontal acquisitions and look for merger opportunities.

COVID-19 pandemic at the beginning of 2020 brought new threats to the auto-industry. Auto-companies have experienced a sharp drop in demand, bottlenecks in supply chains, and plants' lockdown due to governments' trade and mobility restrictions. Joined adverse effects of these factors have resulted in severe cash flow problems, forcing auto-companies to revise the necessity of capital-intensive R&D projects. Firms that can balance exploration and exploitation, sustain operating efficiency, and develop new capabilities in such a hostile business environment will emerge as high-tech leaders in the industry after the pandemic.

Core Messages

- The automotive industry has been continuously facing disruptions since the Great Recession.
- Technological, social, and economic disruptions have triggered the restructuring of the automotive industry.
- Auto-companies have to balance exploration and exploitation to cope with the disruptions.
- Auto-companies use strategic alliances and acquisitions to achieve exploration and exploitation simultaneously.

References

1. Friedman TL (2007) The world is flat: a brief history of the twenty-first century. Farar, Straus and Giroux, New York
2. Mervyn M (2018) Keeping it real: debunking the deglobalization myth, Brexit and Trump: "lessons" on integration. J Int Trade Law Policy 17(1/2):62–68
3. Ghemawat P, Altman SA (2019) The state of globalization in 2019 and what it means. Harvard Business Review. https://hbr.org/2019/02/the-state-of-globalization-in-2019-and-what-it-means-for-strategists. Accessed 09 July 2020
4. Facundo A, Chancel L, Piketty T, Saez E, Zucman G (2018) The elephant curve of global inequality and growth. AEA Papers Proc 108:103–108
5. Milanović B (2016) Why the global 1% and Asian middle class have gain the most from globalization. Harvard Business Review. https://hbr.org/2016/05/why-the-global-1-andthe-asian-middle-class-have-ained-the-most-fromglobalization. Accessed 09 July 2020
6. Khanna T, Palepu KG (2010) Winning in emerging markets: a road map for strategy and execution. Harvard Business Press, New York
7. Bremmer I (2014) The new rules of globalization. Harv Bus Rev 92(1/2):103–107
8. Motohashi K, Yuan Y (2010) Productivity impact of technology spillover from multinationals to local firms: comparing China's automobile and electronics industries. Res Policy 39 (6):790–798
9. Thompson A, Peteraf M, Gamble J, Strickland A (2015) Crafting and executing strategy: the quest for competitive advantage. McGraw Hill, New York
10. Grundy T (2005) Rethinking and reinventing Michaels Porter's five forces model. Strateg Chang 5(8):213–229

11. Porter M (2008) The five competitive forces that shape strategy. Harv Bus Rev 86(1):79–83
12. Farmer REA (2012) The stock market crash of 2008 caused the great recession: theory and evidence. J Econ Dyn Control 36(5):693–707
13. Suarez FF, Lanzolla G (2007) The role of environmental dynamics in building a first mover advantage theory. Acad Manag Rev 32(2):377–392
14. Schlie E, Yip G (2000) Reginal follows global: strategy mixes in the world automotive industry. Eur J Manag 18(4):343–354
15. Pananond P (2016) Break multinationals: emerging—market multinationals in global value chains. In: Handbook of emerging market multinational corporations. Edvard Elgar, pp 91–110
16. Verbeke A, Kano L (2016) An internalization theory perspective on the global and regional strategies of multinational enterprises. J World Bus 51(1):83–92
17. Marković D, Rakita B, Filipović D (2015) Strategic importance of cross-border acquisitions for emerging market multinationals. In: Neostrategic management: an international perspective on trends and challenges. Springer, pp 189–201
18. Yadong L, Sun J, Lu Wang S (2011) Emerging economy copycats: capability, environment, and strategy. Acad Manag Perspect 25(2):37–56
19. OICA production statistics database (2017) International organization of motor vehicles manufacturers. http://www.oica.net/category/production-statistics/. Accessed 09 July 2020
20. Hollweg M (2007) The genealogy of lean production. J Oper Manag 25(2):420–437
21. Klepper S (2002) The capabilities of new firms and the evolution of the US automobile industry. Ind Corp Chang 11(4):645–666
22. Schulze A, MacDuffie JP, Taube FA (2015) Introduction: knowledge generation and innovation diffusion in global automotive industry—change and stability during turbulent times. Ind Corp Chang 24(3):603–611
23. Rugman A, D'Cruz J (1999) The theory of flagship firm. Eur J Manag 15(4):403–412
24. Perez PM, Sanchez AM (2002) Lean production and technology networks in the Spanish automotive supplier industry. Manag Int Rev 42(3):261–277
25. Akpinar M, Zettinig P (2008) Improving supplier power in the buyer-dominated automotive industry. J Transnational Manag 13(3):342–355
26. OICA Sales statistics database (2019) International organization of motor vehicles manufacturers. http://www.oica.net/category/sales-statistics/. Accessed 09 July 2020
27. Flatters P, Willmott M (2009) Understanding the post–recession consumer. Harv Bus Rev 87 (7):106–112
28. Grigolon L, Leheyda N, Verboven F (2016) Scrapping subsidies during the financial crisis: evidence from Europe. Int J Ind Organ 44:41–59
29. Williamson PJ, Ming Z (2009) Value–for–money strategies for recessionary times. Harv Bus Rev 87(3):66–74
30. Moreno A, Terwiesch C (2015) Pricing and production flexibility: an empirical analysis of the US automotive industry. Manuf Serv Oper Manag 17(4):428–444
31. Rakita B, Madić V, Marković D (2017) Competitive strategies of late followers in auto industry: case study Hyundai-Kia. Industrija 45(1):121–146
32. PWC (2018) Charging ahead! The need to upscale UK electric vehicle charging infrastructure. https://www.pwc.co.uk/power-utilities/assets/electric-vehicle-charging-infrastructure.pdf. Accessed 09 July 2020
33. Fishman E (2016) Cycling as transport. Transp Rev 36(1):1–8
34. Kiessling T, Vlacic, B, Dabić M (2019) Mapping the future of cross-border mergers and acquisitions: a review and research agenda. IEEE transactions on engineering management. https://ieeexplore.ieee.org/document/8946901. Accessed 09 July 2020
35. Christensen MC, Alton R, Rising C, Waldeck A (2011) The big idea: the new M&A playbook. Harv Bus Rev 89(3):49–57
36. Kumar N (2009) How emerging giants are rewriting the rules of M&A. Harv Bus Rev 87 (5):115–121

37. Ryoo J (2020) Partnering integration: getting more value from your international acquisition. Organ Dyn 49:1–4
38. Ray S, Gubbi SR (2009) International acquisitions by Indian firms: implications for research on emerging multinationals. Indian J Ind Relat 45(1):11–26
39. Downes L, Nunes P (2013) Big bang disruption. Harv Bus Rev 91(3):44–56
40. Abbosh O, Nunes P, Savic V, Moore M (2017) The big squeeze: how compression threatens old industries. MIT Sloan Manag Rev 58(4)
41. McKinsey and Company (2016) Automotive revolution—perspective towards 2030. https://www.mckinsey.com/~/media/McKinsey/Industries/Automotive%20and%20Assembly/Our%20Insights/Disruptive%20trends%20that%20will%20transform%20the%20auto%20industry/Auto%202030%20report%20Jan%202016.ashx. Accessed 09 July 2020
42. McKinsey and Company (2019) Mastering new mobility. McKinsey center for future mobility. https://www.mckinsey.com/~/media/McKinsey/Industries/Automotive%20and%20Assembly/Our%20Insights/How%20automakers%20can%20master%20new%20mobility/Mastering-new-mobility-Perspectives-on-navigating-an-uncertain-future.ashx. Accessed 09 July 2020
43. EV Volumes the electric vehicle world sales database (2019) EV volumes. https://www.ev-volumes.com/. Accessed 09 July 2020
44. International Energy Agency (2018) Global EV outlook—towards cross-modal electrification. https://www.iea.org/reports/global-ev-outlook-2018. Accessed 09 July 2020
45. Van de Hoed R (2007) Sources of radical technological innovation: the emergence of fuel cell technology in the automotive industry. J Clean Prod 15(11–12):1014–1021
46. Higueras-Castillo E, Kalinic Z, Marinkovic V, Liebana-Cabanillas JF (2020) A mixed analysis of perceptions of electric and hybrid vehicles. Energy Policy 136:1014–1021
47. Jadaan K, Zeater S, Abukhalil Y (2017) Connected vehicles: an innovative transport technology. Procedia Eng 187:641–648
48. PWC (2019) The 2019 strategy and digital auto report. https://www.strategyand.pwc.com/gx/en/insights/2019/digital-auto-report.html. Accessed 09 July 2020
49. Xu J, Liu X (2018) Technology is changing what a premium automotive brand looks like. Harvard business review. https://hbr.org/2018/05/technology-is-changing-what-a-premium-automotive-brand-looks-like. Accessed 09 July 2020
50. Deloitte (2020) Who is going to pay for the future of mobility?. https://www2.deloitte.com/content/dam/insights/us/articles/automotive-news_industry-4-0-in-automotive/6470_whos-going-to-pay/6470_whos-going-to-pay.pdf. Accessed 09 July 2020
51. O'Reilly CA 3rd, Tushman ML (2004) The ambidextrous organization. Harv Bus Rev 82 (4):74–81
52. Hsu CW, Lien YC, Chen H (2013) International ambidexterity and firm performance in small emerging economies. J World Bus 48(1):58–67
53. European Commission Industrial R&D Investment Scoreboard (2019) European Commission. https://iri.jrc.ec.europa.eu/data. Accessed 09 July 2020
54. Vahlne JE, Jonsson A (2017) Ambidexterity as a dynamic capability in the globalization of the multinational business enterprise (MBE): case studies of AB Volvo and IKEA. Int Bus Rev 26(1):57–70
55. Furr NR, Snow DC (2015) Intergenerational hybrids: spillbacks, spillforwards, and adapting to technology discontinuities. Organ Sci 26(2):475–493
56. Wells P, Nieuwenhuis P (2012) Transition failure: understanding continuity in the automotive industry. Technol Forecast Soc Chang 79(9):1681–1692
57. Wiedmann KP, Hennigs N, Schmidt S, Wuestefeld T (2011) Drivers and outcomes of brand heritage: consumers' perception of heritage brands in the automotive industry. J Mark Theory Pract 19(2):205–220
58. Markman GD, Waldron TL (2014) Small entrants and large incumbents: a framework of micro entry. Acad Manag Perspect 28(2):179–197

59. Fosfuri A, Lanzolla G, Suarez FF (2013) Entry-timing strategies: the road ahead. Long Range Plan 46(4–5):300–311
60. Thomas VJ, Maine E (2019) Market entry strategies for electric vehicle startups in the automotive industry–lessons from Tesla motors. J Clean Prod 235:653–663
61. Ferras-Hernández X, Tarrats-Pons E, Arimany-Serrat N (2017) Disruption in the automotive industry: a cambrian moment. Bus Horiz 60(6):855–863
62. McKinsey and Company (2020) The case for an end–end automotive—software platform. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-case-for-an-end-to-end-automotive-software-platform. Accessed 09 July 2020
63. Lockstrom M, Schadel J, Harrison N, Mosser R, Malhotra KM (2010) Antecedents to supplier integration in the automotive industry: a multiple-case study of foreign subsidiaries in China. J Oper Manag 28(3):240–256
64. Langner B, Seidel VP (2009) Collaborative concept development using supplier competitions: insights from the automotive industry. J Eng Tech Manage 26(1–2):1–14
65. Hahn R, Ostertag F, Lehr A, Buttgen M, Benoit S (2020) "I like it but I don't use it": impact of car-sharing business models on usage intentions in the sharing economy. Bus Strategy Environ 29(3):1404–1418
66. Doz Y, Roover B (2019) Responding to digital disruption through alliances. In: Frontiers of strategic alliances research: negotiating, structuring and governing partnerships. Cambridge University Press, pp 101–113
67. Mlađan M, Marković D (2019) Diagrams of power and strategic decision making: the case of strategic alliances in the automotive industry. Manag: J Sustain Bus Manag Solutions Emerg Economies 24(3):21–32
68. Nielsen BB (2007) Determining international strategic alliance performance: a multidimensional approach. Int Bus Rev 16(3):337–361
69. IMF World economic outlook database (2020) International monetary fund. https://www.imf.org/en/Publications/WEO. Accessed 09 July 2020
70. US Bureau of Labor Statistics Payroll Employment Database (2020) US Bureau of labor statistics. https://www.bls.gov/opub/ted/2020/employment-up-2-point-5-million-in-may-2020-but-still-down-19-point-6-million-since-february.htm. Accessed 09 July 2020
71. ACEA Vehicle Registration database (2020) European automobile manufacturers association. https://www.acea.be/press-releases/article/commercial-vehicle-registrations-36.7-five-months-into-2020-44.4-in-may. Accessed 09 July 2020
72. Sturgeon T, Van Biesebroeck J, Gereffi G (2008) Value chains, networks and clusters: reframing the global automotive industry. J Econ Geogr 8(3):297–321
73. Tabrizi B, Lam E, Gerard K, Irvin V (2019) Digital transformation is not about technology. Harvard business review. https://hbr.org/2019/03/digital-transformation-is-not-about-technology. Accessed 09 July 2020

Dušan Marković is an Associate Professor at the Faculty of Economics, Belgrade University, where he also received his Ph. D. in Business. Dušan's Ph.D. thesis studied the competitiveness and economic effects of cross-border acquisitions in Serbia. Professor Markovic teaches the courses on international business and management in the bachelor and master programs at the Faculty of Economics, Belgrade University. He has also taught courses on global competition, international business strategies, innovations, and cross-border acquisitions at the Summer School program, master courses, and MBA courses at EBS Business School, Germany. Professor Markovic also participated in Erasmus's teaching mobility when he taught at DHBW Heilbronn, Germany. He has publications in cross-border acquisitions, emerging markets, international competitive strategies, and the automotive industry. Professor Markovic has rich practical experience in capital and asset valuation, financial due diligence, business, and financial restructuring and has previously worked as a financial advisor for Grant Thornton Serbia.

Veljko Mijušković graduated from the Faculty of Economics, Belgrade, in 2009. He finished his master's studies and got the MSc degree in 2010 at the same faculty. At the end of 2010, he enrolled the Ph.D. studies at the same faculty. He finished the Ph. D. studies in 2017 and obtained his Ph.D. degree. He has been employed since March 2010 at the Faculty of Economics, Belgrade, as a teaching assistant. Starting in October 2017, he has been elected for the position of Assistant professor. He teaches the following subjects: Strategic management, Transport economics, Marketing logistics, Business and Management in a Global context, International logistics, Transport policy and development, Management of logistics, and supply. He has participated in a significant number of scientific national, and international conferences. He is the author of three scientific monographs, co-author of one secondary school textbook and two university textbooks, and a (co)author of numerous papers published in domestic and international journals and monographs. The areas of professional interest are connected with logistics management, supply chain management, and strategic management.

Macroeconomics: From Equilibrium Settings to Multi-agent Worlds

11

Orlando Gomes

John Maynard Keynes [1] famously opined, "If economists could manage to get themselves thought of as humble, competent people on a level with dentists, that would be splendid." He was expressing a hope that the science of macroeconomics would evolve into a useful and routine type of engineering. In this future utopia, avoiding a recession would be as straightforward as filling a cavity.

Mankiw N. G. [2, p. 44].

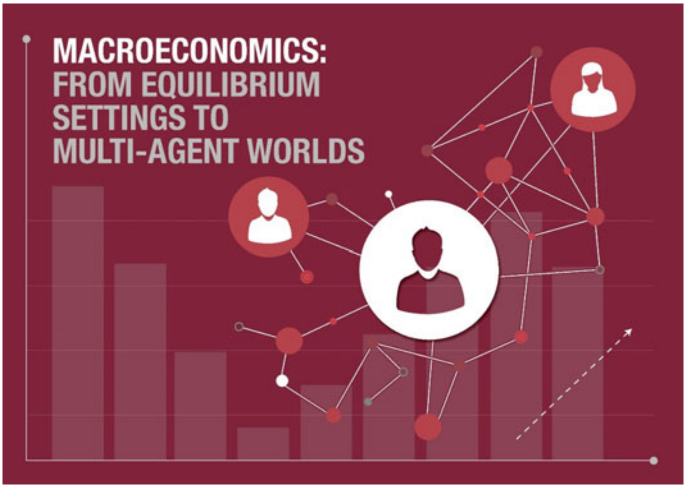
Summary

Macroeconomics studies the performance of the economy as a whole. Trajectories of aggregate variables, as income, consumption, investment, or unemployment, are typically determined, in standard economic theory, through simple mechanic models based on strong assumptions of rationality, efficiency, optimality, and equilibrium. These models have been useful to address many relevant issues and have contributed to a solid understanding of how macro-variables relate and how economic policies might be implemented to increase welfare and enhance growth. Notwithstanding, there is, in such settings, an unrealistic view of how people act: Economic agents are identical in their endowments, preferences, decision-making capabilities, and other features, implying that standard macro-analysis might be pursued under the convention that exists a unique representative agent that solves an infinite horizon optimal control planning problem. As it is frequently pointed out, standard economic theory considers a simple world where a sophisticated agent makes choices. Though, in reality, the opposite occurs: Agents are not

O. Gomes (✉)
Accounting and Business School—Lisbon Polytechnic Institute (ISCAL/IPL) & CEFAGE—ISCAL/IPL Research Center, Lisbon, Portugal
e-mail: omgomes@iscal.ipl.pt

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_11

sophisticated decision-makers, and the world is complex and impossible to process in its entirety by any individual agent. To better understand macro-phenomena, a methodological change is required: Heterogeneity, interaction, bounded rationality, and the use of decision heuristics must be accounted for, in order to further explore the dynamics of collective economic behavior and, therefore, to better understand, predict, and act upon important macro-problems, as recessions, persistent unemployment, or rising inflation.



Credits should be acknowledged to Dr. Filipe Inteiro, Communication Department, ISCAL.

The code of this chapter is 01001101 01100001 01100011 01110010 01101111 01100101 01100011 01101111 01101110 01101111 01101101 01101001 01100011 01110011.

1 Introduction

Research in macroeconomics poses two fundamental challenges. First, one should remark its forward-looking nature. Economics in general, and specifically macroeconomics, deals with future events and their inherent uncertainty, with the way in which such events can be predicted, and with how agents form expectations about the foreseeable evolution of income, investment, employment, inflation, and other aggregate variables. Second, macroeconomics is associated with the study of collective behavior, which is particularly difficult to stylize and model, given the complex dynamic interactions that emerge whenever individual agents engage in contact with one another with the goal of promoting their specific self-interests.

The business transactions that decentralized interaction processes allow for, the institutions that the interplay among agents creates and consolidates, and the policies that public authorities implement to influence aggregate outcomes, are, in fact, pieces of a complex world that constitute the object of study of macroeconomics. As such, it makes sense to associate macroeconomics to the realm of complexity sciences, and one would expect the research in this field of knowledge to rely upon such approaches as agent-based modeling or large-scale simulation studies (techniques that might be, as well, suitable instruments to address the uncertainty issues that were first mentioned).

However, since its inception, macroeconomics has shown resistance regarding the adoption of a multi-agent simulation perspective and, instead, has opted to propose simple toy models built upon basic equilibrium concepts. Partially, this methodological choice has been successful. Many scientific results accomplished in the field are, nowadays, perceived as important tools for policy guidance. Both fiscal policy and monetary policy are, in current days, well supported in scientific knowledge, and scientific breakthroughs have helped many countries in the adoption of successful policies regarding price stability, output and employment stabilization, and poverty alleviation (in this last case, mainly through welfare-state measures).

Notwithstanding, there is a large consensus among macroeconomists that much is yet to be done. Crises as the great recession of 2007–2008 or the most recent negative economic impact of the COVID-19 is difficult to fit into the suit that economists have sewed. Macro-models are elegant representations of how a rigid set of assumptions about the behavior of people and institutions deliver unequivocal results on the performance of the economy. Most of the time, the problem resides on the assumptions, that might be adequate to describe the actions of economic agents in *normal* circumstances to allow for relevant policy guidance, but frequently fail to represent the behavior of people in periods marked by some sort of disruption or distress. The prototype model, under which a representative rational agent maximizes consumption utility, apparently collapses when the premises in which it holds are no longer adequate to describe the world that we live in.

Gradually, the path followed by macroeconomics is shifting. Macroeconomists have realized that to conceive and consolidate a structure of analysis that is simultaneously helpful for a deep understanding of reality and to allow for relevant policy guidance. Some new elements must be brought into the design of the theories. Among these new features, one might highlight heterogeneity (of behavior, endowments, skills, expectations), decentralized interaction, and the use of simple heuristics to replace sophisticated optimal planning.

As the following sections will discuss, some of the above-mentioned features are being introduced in macroeconomic theory in a variety of different modalities. However, a fully integrated approach of heterogeneity-interaction-heuristics in macroeconomics is still missing in the academic literature. The main purpose of this manuscript is to stress the need to develop such a framework, where the parsimony of a single agent taking a limited number of decisions at some initial date for an unspecified time horizon is replaced by a wide array of distinct behaviors, a large number of diverse interaction rules, and a profusion of decision heuristics.

All of this might be combined into complex simulation models that, although somehow voluminous, are capable of representing, characterizing, and predicting observable phenomena with much more accuracy and detail than the standard macro-models. In a way, the fundamental shift consists in putting people into economic models, with their limited knowledge, with their idiosyncrasies, and with their will to benefit from participating in markets and institutions.

The remainder of the chapter is organized as follows. Section 2 reflects on the object of macroeconomics. Section 3 engages in a brief tour through the history of mainstream macroeconomic thought. In Sect. 4, it is emphasized how macro-theory is being reshaped with the help of some important contributions that put in the center of the analysis the idea of multi-agent virtual worlds, where heterogeneous agents endowed with limited knowledge and skills locally interact. Section 5 presents a brief suggestion on how heterogeneity, heuristics, and interactions can be added to macro-frameworks, in the case of a macro-framework dealing with consumption-savings choices. Finally, Sect. 6 concludes.

2 What is Macroeconomics About?

Everyday life decisions require knowledge about how the economy as a whole is structured and how it works. Macroeconomics is the science that studies the overall performance of the economy, by looking at the intertemporal behavior of aggregate variables and how they relate to and influence one another. Such aggregate variables include real values, such as output (GDP—gross domestic product), income (GNI—gross national income), consumption, investment, employment, government expenditures, or taxes; and also nominal variables, including the interest rate, the price level, or the inflation rate.

Understanding macroeconomics requires aggregation, i.e., not only variables, but also agents and markets must be subject to some form of clustering. Concerning economic agents, these are typically aggregated in four different classes: households, firms, financial institutions, and the government. The first three of these classes of agents put together a large array of individuals and entities with distinct endowments, preferences, and objective functions. Government, in turn, is a unique agent that has the capacity and the possibility to take centralized actions with impact over the entire macroeconomic system. The actions taken by governments in an economic context are known as economic policy and they take essentially two forms: fiscal policy (management of tax collection and government expenditures) and monetary policy (manipulation of interest rates with the goal of determining money supply).

In what respects markets, it is typically stipulated that the understanding of the macroeconomy requires conceiving a market for all the goods and services that are transacted (the real market), which is analyzed alongside the money market (supply and demand for money, taking as counterpart other, less liquid, financial assets) and also the markets for production inputs, most predominantly the labor market. The real and money markets constitute the demand side of the macroeconomy and typically receive the greatest attention by economists (who tend to attribute to aggregate demand a particularly important role in the formation of business cycles), while the labor market and the markets for physical capital and technology are the ones that explain the behavior of aggregate supply.

Macroeconomics is essentially about two subjects: the long-term process of wealth accumulation and economic growth; and the short-run aggregate fluctuations that constitute business cycles. Economists typically agree on which are the main determinants of economic growth and on how these might generate a significant and sustained in time increase in per capita income (they include the accumulation of physical inputs—capital and labor—innovation and, at the basis, the consolidation of credible and reliable institutions). They disagree, however, to a large extent, about the causes and propagation mechanisms underlying short-term fluctuations. At this latest level, there is a long-lasting debate between the Keynesian view, grounded on the teachings of [1], and the neoclassical theory, inspired in the work of [3–5], among others.

The main beacon of neoclassical macroeconomists is the faith in the self-regenerating capacity of markets to restore their own equilibrium. If markets are permanently in equilibrium, aggregate fluctuations cannot be associated with any kind of market failure or lack of coordination. In other words, the causes of business cycles must be real in nature and associated with the supply-side of the economy: Technological shocks, disturbances in input markets, or preference changes regarding labor-leisure choices of households, are the candidate triggers of observed fluctuations. On the other hand, Keynesian scholars interpret market imperfections, lack of coordination, information failures, and the sluggish adjustment of prices and wages as the fundamental features underlying business cycles which can, in this case, be characterized as a systematic and continuous departure of markets from their alleged equilibrium position.

In synthesis, despite the disagreement in the interpretation of some phenomena, the object of macroeconomics is well defined: It studies the evolution of aggregate real and nominal variables over time, providing important information for the everyday life of households and firms. Having knowledge on how GDP, employment, investment, inflation, or interest rates evolve is fundamental to make well-informed decisions that allow agents to thrive and prosper. Information on macroeconomic variables and their relationships is also essential for governments to take decisions and implement policies that contribute to increasing prosperity and to a balanced and equitable distribution of the generated wealth.

Looking at the world we live in today, we encounter multiple macroeconomic problems that require attention and action: There are huge per capita income differences between developed and developing countries; income inequality within countries is very high and it is still increasing almost everywhere; the world economy continues to be prone to severe recessions, as the 2007–2008 great recession; the economy reveals its fragility in moments of global turmoil or distress, as the coronavirus pandemic under way; there are still many countries suffering recurrent problems of high unemployment or high inflation, or both.

Macroresearch, as developed in the last decades, was able to build a consistent theory that with relative success addresses some of the issues mentioned in the previous paragraph. However, this theory presents limitations: It cannot explain all observable phenomena and it is often surprised by how reality evolves. Furthermore, it has not provided the policy guidance that many states would desire to solve or soften the most pressing problems concerning poverty, unemployment, or increased cost of living. In the following section, a brief tour of the success and failures of mainstream macro-theory is undertaken, paving the way for the discussion that follows on alternative approaches to characterize and explain the macroeconomy.

3 Mainstream Macro-Theory

Despite the well-known Keynesian—neoclassical controversy in macroeconomics, there is a relatively wide consensus regarding the techniques of analysis to employ to approach macro-events and macro-relations. This consensus is built upon the use of representative agent models and intertemporal optimization problems. Because agents form expectations and make decisions rationally, they adopt a similar behavior, and therefore the macroeconomy might be scrutinized through the analysis of the behavior of a median or representative agent. Heterogeneity plays no role in orthodox macroeconomic thinking, although heterogeneity is, in fact, the element that, in reality, underlies the richness associated with collective actions.

In the paragraphs that follow, we undertake a brief tour over the most emblematic macroeconomic theories produced over the last few decades.[1] In them, there is a pervasive feature, which is precisely the notions of representative household and representative firm (or a single representative agent), accompanied by a planning behavior that frequently consists of maximizing or minimizing an objective function over a long or even infinite horizon. We go through growth theory, households' consumer choice, dynamic stochastic general equilibrium models, new Keynesian theory, and models of matching unemployment.

In growth theory, the benchmark model that explains intertemporal consumption choices is the optimal control problem designed by Ramsey [7] and further elaborated by [8, 9]. In this model, a forward-looking representative agent plans, at a given initial date, the future trajectory of consumption that serves the goal of maximizing consumption utility over an infinite horizon. Basically, the decision to make is how much to consume on every date, which is the same as saying that the decision is about how much to save and spend along the assumed lifecycle.

In growth models, the decision of the representative agent is subject to a series of constraints, mainly related to the accumulation of material inputs (physical capital and human capital), but also with innovation and the generation of new ideas, processes that are fundamental to enhance the total factor productivity with which material inputs contribute to production. These prototypical models allow for relevant insights into the growth process. Overall, they indicate that long-term

[1] See [6] for a detailed and comprehensive study of the most prominent theories developed under the auspices of modern macroeconomics.

sustained growth requires more than capital accumulation, which is subject to decreasing marginal returns. Innovation and education are fundamental drivers of growth in the long term and, therefore, they must play a central role in any theory of economic growth.

Benchmark growth theory also offers a good platform to discuss convergence and divergence of income across nations. While a simplistic view based on the diminishing marginal returns approach may point to convergence (diminishing returns affect more intensely those economies in an advanced capital accumulation stage), the empirical experience reveals a strong variety of outcomes: Some countries are effectively converging, while other economies diverge, what points to a possible polarizing process regarding the pace of growth in our world. This diversity of growth processes is well accommodated by the simple paradigm of the representative agent who maximizes utility subject to wide economy constraints, as long as one accounts for a series of heterogeneous exogenous factors (i.e., different rates of population growth, different savings rates, different rates of technological progress, and different rates of capital depreciation).

Unlike growth models, macro-theories focused on the short-run, which are essentially related to the explanation of causes and consequences of cyclical fluctuations, tend to consider uncertainty and, thus, a stochastic component. This stochastic component is essentially associated with the formation of expectations. Since [10], the main assumption underlying the formation of expectations by economic agents is the rational expectations hypothesis. If agents formulate expectations in the same rational way, the immediate corollary is that agents are identical and, also in this context, the economy might be analyzed under the perspective of the representative agent.

Again, the fundamental piece of the theoretical reasoning is the intertemporal consumption utility framework. Given the budget constraint of the household, she will maximize the expected utility. The result is basically such that the agent desires to smooth consumption over the life cycle, an outcome first highlighted by [11, 12] in their life cycle and permanent income theories.

Putting together the consumption optimization framework with a capital accumulation constraint and adding to this setup labor-leisure choices, which allows to endogenize labor supply, it is possible to transform the typical optimal growth model into a dynamic stochastic general equilibrium (DSGE) model. The popular DSGE models are the basic framework through which neoclassical economists approach business cycles. Following the work of [5], which has initiated the Real Business Cycle (RBC) theory, one may add to such framework the possibility of exogenous shocks over technology. This allows us explaining how eventual stochastic innovation processes trigger fluctuations, i.e., periods of expansion followed by periods of recession. The popularity of RBC models came from the fact that a relatively simple model, based on the foundations of rational decision and rational behavior, could replicate with some degree of accuracy the stylized facts on short-term business cycles.

The neoclassical theory of business cycles, i.e., the RBC model, is subject to criticism, namely because the mechanism of propagation of fluctuations in this setting is essentially labor supply. If one recognizes that the wage elasticity of labor supply is low, this signifies that the underlying causes of business cycles must be other than supply-side shocks and their effect on labor-leisure choices. Keynesian authors have adopted a different view, namely a perspective in which the main underlying causes of business cycles are essentially attached to the rigidity of prices and wages and how this rigidity provokes inertia whenever a policy shock or a preference shock takes place.

The staggered prices model of [13] and the menu costs framework proposed by Mankiw [14] are two good examples of how the Keynesian view of imperfect markets and semi-rational behavior could be adapted in order to integrate the standard framework of analysis, which is still the representative agent DSGE framework.

One of the issues that mainstream macro-theory had difficulty in dealing with is unemployment. After all, if markets are permanently in equilibrium or even if they are just sluggish to adapt given the rigidity of prices and wages, unemployment would not be a fundamental problem, but only a transitory inconvenience. In any circumstance, the labor market, as any other market, would exhibit a tendency to adjust to the equilibrium position, an equilibrium that is Pareto efficient and, therefore, where unemployment cannot persist. However, reality shows that unemployment tends to be persistently high even in developed economies where institutions work well, information circulates freely, and agents are also free to react to incentives.

The most celebrated model of unemployment in economics is the search and matching model proposed by [15, 16]. The search and matching model takes not one representative agent but two agents, households and firms, that solve distinct optimization problems. The optimal control problem of firms consists in maximizing profits given their estimated revenues and the costs associated with paying wages and hiring new employees; in turn, households maximize their income, which comes from wages when employed and from some type of public assistance when unemployed. The solution of each of the above problems gives place to two different wages: The wage employees optimally expect to receive, and the wage employers optimally want to pay. There is no reason for these to coincide, and therefore, only through negotiation the two agents might reach a mutually beneficial outcome, which culminates with the worker occupying the vacancy offered by the employer.

The search and matching model introduces a minimal degree of heterogeneity. Unlike other macro-models, there is no merging of households and firms into a single agent with a common goal and common constraints. Nevertheless, the analysis continues to be based on an extreme degree of aggregation, where the interaction between agents with distinct preferences, endowments, and expectations dominates.

The frameworks that mainstream macro-theory was able to construct are relevant to understand important phenomena regarding the link between the behavior of agents and short-term and long-term performance of the economy and of the main economic aggregates. It has been successful in many respects, namely concerning the way it can inform political decision-makers. Through fiscal policy and monetary policy, governments and central banks have been able to tackle, with more or less success, the problems of recessions, unemployment, income inequality, price stability, and interest rate stability. We would not have growing economies with good living standards all around the world without following the powerful teachings that come from decades of a thorough analysis and modeling of the macroeconomic system.

Nevertheless, much is still to be done in what concerns the understanding of the functioning of the economy and the implementation of successful public policies. Additional steps must be taken in order to arrive in a more comprehensive theory capable of taking into account the idiosyncrasies of individual agents and the way the interaction among heterogeneous agents produces unique emergent phenomena. Until now, macroeconomics has put itself in a position where it can be contested on the grounds of the fallacy of composition: In the economy, as with regard to any other social or natural object, the whole is in no way the sum of the parts. The whole is the outcome of the complex interaction among simple but distinct individual pieces.

4 The Way Forward: Non-Optimal Decision-Making, Heterogeneity, and Interaction

The representative agent model frontally clashes with the observable evidence. Benartzi and Thaler [17] highlight that associated with the explicit representative agent assumption, there is a pair of two additional implicit assumptions that are in clear contrast or conflict with reality: First, in such a setting, agents cannot face any cognitive constraints and, second, agents must have the required willpower to fully and successfully execute the optimal plans they formulate.

As such, macro-theories based on the assumption of optimality and rational behavior suggest a planning ability, a cognitive capacity, and an unshakable will that represent a *homo-economicus* which can only be found in books written by economists. As [18] put it, it would be necessary for a supercomputer and a Ph.D. degree in economics for common households to be able to formulate and solve the kind of plans economists consider in their macroeconomic frameworks.

Frequently, household decisions are based on heuristics or rules-of-thumb, which do not conduct to the optimal outcome, but is feasible to formulate under the cognitive constraints faced by human beings. In the real world, populated by *homo-sapiens* and not by *homo-economicus*, individual agents face a multiplicity of constraints and obstacles when collecting and processing the relevant information required for decision-making. Such obstacles justify the reason why computation of optimal solutions is, frequently, from the start, replaced by the adoption of simple idiosyncratic choice rules.

Explaining human behavior, in the context of science, through heuristics, might be controversial, and certainly a less elegant solution than the one provided by the rationality approach. Finding heuristics that are pervasive in use and that unambiguously capture human behavior is a messy task, as highlighted by Haldane and Turrell [19], who indicate that for such a trivial subject as the formulation of a simple household consumption rule, the relevant literature has proposed an endless list of possible candidates. Nevertheless, this is the work of science: going through messy processes to arrive at frameworks of analysis that constitute important breakthroughs, and in what concerns decision heuristics this has been accomplished.

In [20–22], heuristics are classified as fast and frugal rules that make the best possible use of the available information given the constraints of time and knowledge that individuals face while conducting their deliberation processes. On many occasions, making well-thought, deep and careful decisions is not just possible. These authors go even further by claiming that on many occasions less processing might imply better inferences. This is the less-is-more principle, which has been used to justify that heuristics are not necessarily second-best approximations to optimal planning. Rules-of-thumb may outperform sophisticated forecasting tools-based onstrict rationality principles.

The use of simple decision rules in macrosettings, as highlighted in [23–29], is a possible path to replace the orthodoxy of optimal behavior and optimal forecasting. They are an important piece of the novel macroeconomic paradigm, but not the single one. While the optimal behavior assumption automatically implies that all agents behave identically, decision heuristics may diverge across individuals, opening the door to another fundamental feature underlying the new paradigm of analysis: heterogeneity.

The main problem with standard macroeconomics is that it falls in the so-called fallacy of composition, in which the whole is interpreted as being the mere sum of the parts. This overlooks the richness that arises once we consider the heterogeneity of preferences, endowments, capabilities, and expectations that lead, through interaction, to the prevalence of emergent phenomena. Even within the current orthodoxy, though, many macro-models are evolving in the direction of including elements of heterogeneity and interaction, as we remark in what follows.

In two relevant studies, [30, 31] have proposed macro-models with two different classes of agents: the optimal planners or Ricardian consumers, and those who live hand-to-mouth (the non-Ricardian consumers, who ignore intertemporal trade-offs). These models are suitable to address monetary and fiscal policy implications in economies where agents do not share the exact same behavior. In [32] such type of analysis is extended, assuming instead of a two-agent framework, a setting where a multiplicity of different household consumption and savings behavior is allowed for. The Kaplan model received the designation of HANK (heterogeneous agents New Keynesian) model, to contrast with the TANK (two-agent New Keynesian) model and the RANK (representative agent New Keynesian) model.

Also, in [33] a standard optimal growth model is adapted in order to contemplate agent heterogeneity, decentralized interaction, and non-optimal behavior. This model constitutes a guide on how a conventional equilibrium growth model can be transformed into a multi-agent setting, with decentralized interaction among boundedly rational heterogeneous agents.

Under this more reasonable interpretation of human behavior, where agents are allowed to follow diverse strategies regarding decision-making and where such strategies do not have to coincide with optimal rational planning, and instantaneous relevant corollary emerges: psychological profiles matter for macroeconomics. This is true, for instance, in what concerns the fundamental choice between consuming now or later, i.e., the fundamental choice between consumption and savings.

The theory of rational behavior indicates that agents will want to smooth consumption, and therefore to save when their income is high to spend above income when their income is low. However, evidence shows that there is sentiment heterogeneity regarding savings (see [34]): There is a multiplicity of psychological factors influencing savings behavior (personality traits, self-control, regulatory focus, degree of optimism) that make some people be savers while others are, in a larger or smaller extent, spenders. At the end of the day, not all psychological profiles generate consumption smoothing; some households continue to save even when future consumption is more than guaranteed, while other households adopt a hand-to-mouth behavior with no concern about the future. We will get back to consumption-savings heterogeneity later in Sect. 5.

Synthesizing the discussion so far, one might say that there is an important evolution under way in economic thinking: The standard optimal planning framework has been refined or even replaced with behavioral features, where the actions of agents are modeled in a way that is more compatible with actual and observable decision-making processes. As remarked, for instance, by [35, 36], the world is too complex to be fully understood even by agents with an influential role in the economy, like governments or large corporations. In such a context, agents adopt different strategies and courses of action and promote interaction to attain their goals; in such a context, as well, the use of heuristics in the decision-making process is not a symptom of lack of rationality, rather a way to cope with the complexity of the world.

Typical economic models, where economic agents make use of a supernatural capacity to collect and process information to solve intertemporal optimization problems, are, therefore, receding in their scientific status, given the recognition that they are not truly reflexive of the complexity of the world and of the dynamics that emerge once interaction among different people is considered. In [37], it is pointed out that the economy and its agents are, indeed, complex systems. They are composed of a series of elements that interact in non-trivial ways to generate unique and unrepeatable outcomes.

In complex systems, spontaneous orders emerge and evolve without the intervention of any global controller or central planner. This results in an aggregate outcome that is path-dependent (events are historically determined) and that contrasts with the economic orthodoxy built upon the concepts of equilibrium, efficiency, optimization, and rationality. Under the complexity view, matters most the actual behavior of people than the conceptual framework associated with ideal optimal behavior. The "*as is*" gains predominance over the "*as if"* as a scientific premise.

Adopting a complexity perspective implies understanding that there is a permanent interaction between the micro- and the macro-levels. Micro-decisions and micro-actions shape the macro-dynamics, but these also influence the posterior behavior of the individual entities, and so forth. A successful macro-model, capable of giving a robust and intuitive explanation of reality, must account for the mentioned elements of complexity.

[38, 39] discusses the notion of the artificial world in the context of the study of the macroeconomy. The author's interpretation of artificial world consists of a complex system based on an emergent hierarchical organization. Agents interact within a system composed of different hierarchical levels, and the hierarchical levels are formed and eventually disappear through the different types of relations established among agents. Every artificial world conceived under this perspective will share three elements: a set of entities at the micro-level, an environment where to interact, and a dynamic interaction process. The attributes of the microentities and the shape of the macroenvironment are constantly being reshaped under the designed interaction process.

Artificial worlds, generated in the way characterized above or in other similar ways, allow for the identification of emergent properties and macro-patterns. The dynamics of an artificial world are, in fact, an evolutionary process where mechanisms of replication, selection, and variation lead to an infinity of possible outcomes.

The new path for macroeconomic analysis is also strongly associated with the agent-based literature. According to [24], an agent-based approach is a bottom-up approach in which heterogeneous agents adopt simple rules to interact in a complex world. Furthermore, the resulting aggregate phenomena are emergent phenomena that arise from the apparently uncoordinated actions of a large number of individuals, who are not guided by any pre-determined equilibrium (differently from what happens in standard macro-theory).

The agent-based approach is a part of a wider trend in science in general, and in economics in particular, regarding the replacement of the mechanic equilibrium perspective, by another perspective that is more procedural, algorithmic, and simulation-based [40]. This is the complexity view, which sees socio-economic systems as complex systems that might be modeled through the simulation of artificial worlds. The ideas and notions of agent heterogeneity, bounded rationality, network connectivity, emergence, and out-of-equilibrium dynamics are the key concepts of this new paradigm, as characterized in [41–46].

The main obstacle associated with the implementation of a complexity view in macroeconomics concerns the many degrees of freedom that such an approach allows for. There is still, in the current state of affairs, too much arbitrariness in the definition of assumptions, in the selection of heuristics, in the formalization of sequences and timing of events, and in the design of rules of interaction. Conceiving a well-structured complex system capable of capturing the subtleties of the interaction of agents in a large-scale economy is a difficult task, which requires a sensible balance between simplicity and comprehensiveness. This is not easy to accomplish, but it is the endeavor that science must persist in trying to accomplish.

5 Heuristics and Heterogeneity: An Example

To briefly illustrate the path macroeconomics might follow once the notions of heterogeneity, local interaction, and rule-of-thumb behavior are fully assimilated, I recover the consumption decision rule proposed in [47]. This rule replaces the typical optimal household behavior with a static rule, in which agents might adopt different attitudes toward consumption and savings. Taking the terminology employed in [47], individual households might be ants (savers) or grasshoppers (spenders) in different degrees.

Agents are indexed by $x \in \mathbb{R}$. If $x < 0$, then the agent is an ant, i.e., an agent with a given propensity to save, which increases with the absolute value of the index; if $x > 0$, then the agent might be classified as a grasshopper, i.e., an agent that spends in consumption all of its income or an amount larger than current income. The design of the consumption rule requires the definition of the desired consumption threshold, such that

$$\overline{C}_t = e^x Y_t \tag{1}$$

In Eq. (1), $\overline{C}_t$ is the agent's desired consumption which equals the agent's income level, Y_t, multiplied by a term that reflects the household's psychological profile. Whenever $x < 0$, the consumption level desired by the agent remains below the received income (this is the definition of an ant, in the current context). In the case $x > 0$, desired consumption exceeds income, what is the characterization of the behavior of a grasshopper.

To simplify the presentation, assume that the agent's income corresponds solely to the returns from wealth accumulation (i.e., labor returns are ignored). Considering that wealth, $A_t \geq 0$, is remunerated at a rate of return r, the income level at period t is $Y_t = rA_t$.

Given the above definitions, the proposed consumption rule is

$$C_t = \begin{cases} A_t + Y_t \; if \; Y_t \leq \overline{C}_t \text{ and } A_t + Y_t \leq \overline{C}_t \\ \overline{C}_t \; if \; Y_t \leq \overline{C}_t \text{ and } A_t + Y_t > \overline{C}_t \\ \overline{C}_t + \zeta\left(Y_t - \overline{C}_t\right) \text{ if } Y_t > \overline{C}_t \end{cases} \tag{2}$$

Wealth accumulation obeys the following simple dynamic process,

$$A_{t+1} - A_t = Y_t - C_t \Leftrightarrow A_{t+1} = (1+r)A_t - C_t, \;\; A_0 \text{ given} \tag{3}$$

where $r \geq 0$ is the real interest rate. Wealth increases with asset returns and decreases with consumption.

Equation (2) has the following interpretation: When households are grasshoppers at an extreme point in which they would want to consume more than the accumulated wealth plus current income, they will in fact consume the aggregate wealth level plus income they possess in that specific time period. If the household is a grasshopper but her desired consumption remains below the accumulated wealth and income levels, then consumption will coincide with desired consumption. Whenever the agent desires to consume less than current income, the agent will save a share $1-\zeta \in (0,1)$ of the difference between income and desired consumption.

Consumption heuristic (2) might be rewritten under the following form,

$$C_t = \begin{cases} \frac{1+r}{r} Y_t \text{ if } x \geq \ln\left(\frac{1+r}{r}\right) \\ e^x Y_t \text{ if } 0 \leq x < \ln\left(\frac{1+r}{r}\right) \\ [e^x + \zeta(1-e^x)] Y_t \text{ if } x < 0 \end{cases} \tag{4}$$

Equation (4) makes it explicit the existence of two relevant thresholds: An agent for whom $x \geq \ln\left(\frac{1+r}{r}\right)$ will consume all her available wealth; an agent for whom $0 \leq x < \ln\left(\frac{1+r}{r}\right)$ consumes a multiple of her income; a household such that $x < 0$ consumes a share $e^x + \zeta(1-e^x)$ of the corresponding income. The savings rate of this last type of agent is $(1-\zeta)(1-e^x)$. Note that the heterogeneity in this rule is present in the level of parameter x which indicates the extent in which an agent is an ant or, alternatively, a grasshopper.

Given the established relations, it is straightforward to compute the growth rate of consumption in any of the three considered circumstances. Computation allows deriving the following outcome,

$$\frac{C_{t+1} - C_t}{C_t} = \begin{cases} 0 \text{ if } x \geq \ln\left(\frac{1+r}{r}\right) \\ r(1-e^x) \text{ if } 0 \leq x < \ln\left(\frac{1+r}{r}\right) \\ r(1-\zeta)(1-e^x) \text{ if } x < 0 \end{cases} \tag{5}$$

In the scenarios given by the first two segments of the consumption rule, consumption falls to zero. In the first case, this occurs instantly, as the household consumes all the available wealth at $t = 0$ and, thus, no more wealth is available to generate new income. In the second case, accumulation of wealth continues after the first period, as long as $x < ln\left(\frac{1+r}{r}\right)$, but the growth rate of consumption will be negative and, therefore, in this case, consumption also falls to zero. The only scenario in which consumption continues to grow in the long-term is the third one, i.e., the scenario in which agents save part of the income they access to in each period. This positive and constant growth rate means that the third scenario is the only one feasible if the agent intends to perpetuate consumption over time.

The long-run scenario contrasts with the short-run outcome. Observe that, at $t = 0$, the following inequality holds,

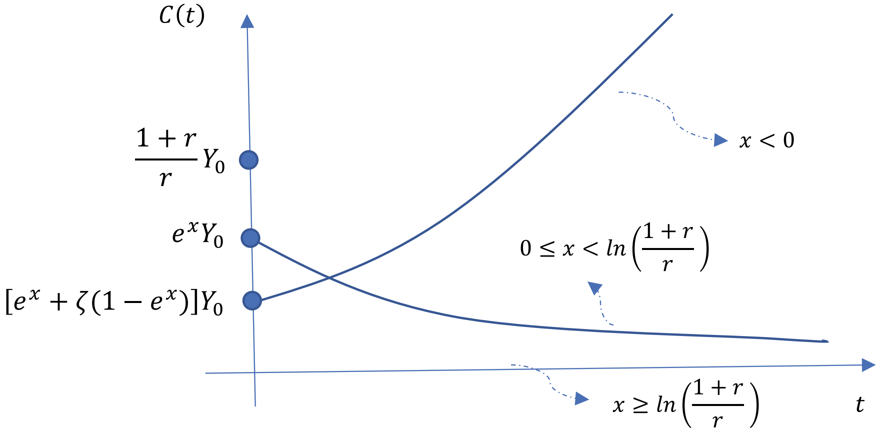


Fig. 1 Consumption paths

$$\left.\frac{1+r}{r}Y_0\right|_{x \geq \ln\left(\frac{1+r}{r}\right)} > \mathrm{e}^x Y_0|_{0 \leq x < \ln\left(\frac{1+r}{r}\right)} > [\mathrm{e}^x + \zeta(1-\mathrm{e}^x)]Y_0|_{x<0} \tag{6}$$

Therefore, the best solution in the long run, the one that allows consumption to continue to grow indefinitely, is the one that penalizes most the agent in the current initial period. Figure 1 depicts this outcome, for three different values of x, each one representing one of the three possible cases.

The above reasoning, synthesized in the graphic of Fig. 1, allows for two important considerations in the context of the macroeconomics change of paradigm: First, agents do not necessarily formulate sophisticated plans; for instance, regarding consumption, as is the case, agents may simply establish a benchmark of how much to spend given their prospective income. Second, by adopting a simple rule, one also opens the door to eventual heterogeneity of behavior; the specific assumed rule allows for agents to choose if they are spenders or savers to a certain extent. Not all agents are identical and some will give preference to the short-run relatively to the long-run or the opposite. Ants and grasshoppers certainly coexist in the economy given different degrees of impatience and care about future outcomes.

A third element, which can be added to the analysis, is the possibility of changing behavior through local interaction. Grasshoppers in contact with ants may perceive the advantages of saving for the future, in order to perpetuate consumption growth; or the opposite: Ants in contact with grasshoppers may be convinced of the short-run advantages of decreasing the absolute value of their x or even of putting it in the positive side. If this type of local interaction is systematic, leading to constant changes of adopted behavior, nonlinear aggregate outcomes will be found, as agents are constantly changing their strategies in order to balance the short-run/long-run trade-off.

6 Conclusion

Modeling the macroeconomy is not, in any circumstance, an easy task. There are too many variables that are intertwined in many ways. Moreover, each aggregate variable represents a large set of features that lose their individuality when placed under a single aggregate. Macroeconomics has made important progress along the last few decades in presenting relevant explanations for phenomena concerning the consumption-savings decisions of households, the investment decisions of firms, the policy choices made by governments and central banks, and the evolution of relevant macro-indicators as the unemployment rate or the inflation rate.

Nevertheless, and despite the elegance and robustness of the proposed models, researchers, policymakers, and the general public tend to feel that macroeconomics should go further and deeper in its explanatory capacity. The main failure seems to be related to the simplifying assumption regarding the notion of the representative agent. No meaningful explanation of macro-events is possible without understanding the dynamics of collective action and the way aggregate outcomes emerge, not from the simple averaging of individual behaviors but instead from complex decentralized interaction processes.

Therefore, macroeconomics needs a paradigm shift. Without neglecting all the scientific progress that has been accomplished, it is now clear in the mind of the economists that the science can only progress, in the benefit of a sounder understanding of actual phenomena, if the equilibrium-based analysis is replaced by a procedural, algorithmic and organic interpretation of reality. This new interpretation of reality necessarily relies on what complexity science has to offer, and in three pillars that are essential to construct a macro-complex theory. These are heterogeneity, decentralized interaction, and decision-making under simple rules. Developing macroeconomic models based on these three vectors will be a fundamental step to gain better knowledge on the collective processes that generate the observable aggregate phenomena.

After reviewing the accomplishments of macroeconomic theory, and emphasizing the paths for a new paradigm, the manuscript has proposed a simple piece to fit the puzzle of the new approach. When making the important decision of consuming now or saving to consume later, there are two elements of primordial relevance: First, the fact that such decision is typically made under a certain view of the world that make people more or less propensity to save; and, second, the fact that such heterogeneous agents are unsophisticated decision-makers that simply choose to consume relatively more or less to the received income; no intricate optimization problem is effectively solved by those who have to equate how much to spend each week, month, or year, given their current and expected income levels.

As with regard to consumption decisions, in any area of macroeconomics it is necessary to transform the current frameworks of analysis into worlds populated by a multitude of agents with different characteristics, endowments, behaviors, and expectations. The analysis of the interplay between individual agents in more or less structured networks of relations will certainly result in a much richer understanding of the relevant phenomena and a better capacity for public agents to act upon reality.

Core Messages

- Macroeconomics has made important progress over the last decades;
- Conventional macro-models are based on the notions of equilibrium, rational behavior, and optimal planning;
- The science of macroeconomics needs a paradigm shift;
- The equilibrium-based analysis must be replaced by a complexity approach;
- Three pillars are essential to construct a macro-complex theory: heterogeneity, decentralized interaction, and decision-making based on simple heuristics;
- Consumption-savings decisions are better understood under the heterogeneity-interaction-heuristics framework.

References

1. Keynes JM (1936) The general theory of employment, interest, and money, 2018th edn. Palgrave Macmillan, Cham, Switzerland
2. Mankiw NG (2006) The macroeconomist as scientist and engineer. J Econ Perspect 20:29–46
3. Friedman M (1968) The role of monetary policy. Am Econ Rev 58:1–17
4. Lucas RE (1977) Understanding business cycles. Carnegie Rochester Conf Ser Public Policy 5:7–21
5. Kydland FE, Prescott EC (1982) Time to build and aggregate fluctuations. Econometrica 50:1345–1370
6. Alogoskoufis G (2019) Dynamic macroeconomics. MIT Press, Cambridge, MA
7. Ramsey F (1928) A mathematical theory of saving. Econ J 38:543–559
8. Cass D (1965) Optimum growth in an aggregative model of capital accumulation. Rev Econ Stud 32:233–240
9. Koopmans TC (1965) On the concept of optimal economic growth. In: The econometric approach to development planning. North Holland, Amsterdam
10. Lucas RE (1972) Expectations and the neutrality of money. J Econ Theory 4:103–124
11. Modigliani F, Brumberg R (1954) Utility analysis and the consumption function: an interpretation of cross section data. In: Kurihara K (ed) Post Keynesian economics. Rutgers University Press, New Brunswick, NJ, pp 388–436
12. Friedman M (1957) A theory of the consumption function. Princeton University Press, Princeton, NJ
13. Calvo GA (1983) Staggered prices in a utility maximizing framework. J Monet Econ 12:383–398
14. Mankiw NG (1985) Small menu costs and large business cycles: a macroeconomic model of monopoly. Quart J Econ 100:529–537

15. Pissarides CA (1985) Short run dynamics of unemployment, vacancies and real wages. Am Econ Rev 75:676–690
16. Mortensen DT (1986) Job search and labor market analysis. In: Ashenfelter O, Layard PRG (eds) Handbook of labor economics, vol 2. Elsevier, Amsterdam, pp 849–919
17. Benartzi S, Thaler R (2007) Heuristics and biases in retirement savings behavior. J Econ Perspect 21:81–104
18. Allen TW, Carroll CD (2001) Individual learning about consumption. Macroecon Dyn 5:255–271
19. Haldane AG, Turrell AE (2019) Drawing on different disciplines: macroeconomic agent-based models. J Evol Econ 29:39–66
20. Gigerenzer G, Brighton H (2009) Homo heuristicus: why biased minds make better inferences. Top Cogn Sci 1:107–143
21. Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Annu Rev Psychol 62:451–482
22. Dosi G, Napoletano M, Roventini A, Stiglitz JE, Treibich T (2017) Rational heuristics? Expectations and behaviors in evolving economies with heterogeneous interacting agents. LEM papers series 2017/31, Sant'Anna School of advanced Studies, Pisa, Italy
23. Dosi G, Fagiolo G, Roventini A (2010) Schumpeter meeting Keynes: a policy-friendly model of endogenous growth and business cycles. J Econ Dyn Control 34:1748–1767
24. Dosi G, Fagiolo G, Napoletano M, Roventini A (2013) Income distribution, credit and fiscal policies in an agent-based keynesian model. J Econ Dyn Control 37:1598–1625
25. Dosi G, Fagiolo G, Napoletano M, Roventini A, Treibich T (2015) Fiscal and monetary policies in complex evolving economies. J Econ Dyn Control 52:166–189
26. Dosi G, Roventini A (2019) More is different … and complex! The case for agent-based macroeconomics. J Evol Econ 29:1–37
27. Dosi G, Roventini A, Russo E (2019) Endogenous growth and global divergence in a multi-country agent-based model. J Econ Dyn Control 101:101–129
28. Dawid H, Harting P, Neugart M (2014) Economic convergence: policy implications from a heterogeneous agent model. J Econ Dyn Control 44:54–80
29. Dawid H, Harting P, Neugart M (2018) Cohesion policy and inequality dynamics: insights from a heterogeneous agents macroeconomic model. J Econ Behav Organ 150:220–255
30. Gali J, López-Salido JD, Vallés J (2004) Rule-of-thumb consumers and the design of interest rate rules. J Money Credit Bank 36:739–763
31. Gali J, López-Salido JD, Vallés J (2007) Understanding the effects of government spending on consumption. J Eur Econ Assoc 5:227–270
32. Kaplan G, Moll B, Violante GL (2018) Monetary policy according to HANK. Am Econ Rev 108:697–743
33. Gomes O (2020) From conventional equilibrium models to multi-agent virtual worlds: a prototype economic growth example. Nonlinear Dyn Psychol Life Sci 24:233–260
34. Gerhard P, Gladstone JJ, Hoffmann AOI (2018) Psychological characteristics and household savings behavior: the importance of accounting for latent heterogeneity. J Econ Behav Organ 148:66–82
35. De Grauwe P (2011) Animal spirits and monetary policy. Econ Theor 47:423–457
36. Gabaix X (2014) A sparsity-based model of bounded rationality. Quart J Econ 129:1661–1710
37. Gomes O (2019) Simulation games in economics and business: building artificial worlds for flesh and blood players. In: Gomes O, Gubareva M (eds) Contributions on applied business research and simulation studies. Nova Science Publishers, Inc., Hauppauge, New York, Chapter 8
38. Lane DA (1993) Artificial worlds and economics, part I. J Evol Econ 3:89–107
39. Lane DA (1993) Artificial worlds and economics, part II. J Evol Econ 3:177–197
40. Arthur WB (2013) Complexity economics: a different framework for economic thought. Santa Fe Institute working paper no 2013-04-012

41. Delli GD, Gaffeo E, Gallegati M (2010) Complex agent-based macroeconomics: a research agenda for a new paradigm. J Econ Interact Coord 5:111–135
42. Fagiolo G, Roventini A (2017) Macroeconomic policy in DSGE and agent-based models redux: new developments and challenges ahead. J Artif Soc Soc Simul 20(1)
43. Holt RPF, Rosser JB, Colander D (2010) The complexity era in economics. Middlebury College Economics discussion paper no 10-01
44. Le Baron B, Tesfatsion L (2008) Modeling macroeconomies as open-ended dynamic systems of interacting agents. Am Econ Rev 98:246–250
45. Rosser JB, Rosser MV (2015) Complexity and behavioral economics. Nonlinear Dyn Psychol Life Sci 19:201–226
46. Stiglitz JE, Gallegati M (2011) Heterogeneous interacting agent models for understanding monetary economies. East Econ J 37:6–12
47. Gomes O (2020) Growth theory under heterogeneous heuristic behavior. Forthcoming. J Evol Econ 31:533–571

Orlando Gomes is Professor of Economics at Lisbon Accounting and Business School of the Lisbon Polytechnic Institute (ISCAL-IPL, Portugal). He holds a Ph.D. in Economics and has published more than 100 scientific papers in international journals in Economics and related areas. Some recent publications can be found in the following journals: Journal of Evolutionary Economics, Journal of Economic Coordination and Interaction, Macroeconomic Dynamics, Nonlinear Dynamics in Psychology and Life Sciences, Physica A, Economic Modelling, and Metroeconomica. His research interests cover areas such as macroeconomics, economic growth, economic policy, economic theory, nonlinear dynamics, agent-based modeling, and complexity.

12 A Multi-level Perspective to Biosimilars Development: Pathways Towards Incremental Innovation in the Health Bioeconomy

Ruth Oriama, Robert Mudida, and Thierry Burger-Helmchen

> *We see extraordinary changes sweeping through the pharmaceutical industry. There is a well-established pattern by which these changes arise. They always have loomed as threats to some, and opportunities to others—and whether it is one or the other is a choice.*
>
> Clayton M. Christensen

Summary

Biosimilar development is one of the extraordinary changes sweeping through the pharmaceutical industry, providing an opportunity for incremental innovation in the health bioeconomy domain. The dawn of the New Biology, global knowledge diffusion that inspires reverse innovation, and an increased political commitment to the real subsumption of nature through the bioeconomy characterize biosimilar development. These advancements come in the backdrop of high health-care costs and an ambition to achieve universal health care for all. This chapter argues that considerations of the multi-level perspective are necessary for a health bioeconomy transformation. Emerging borderless themes

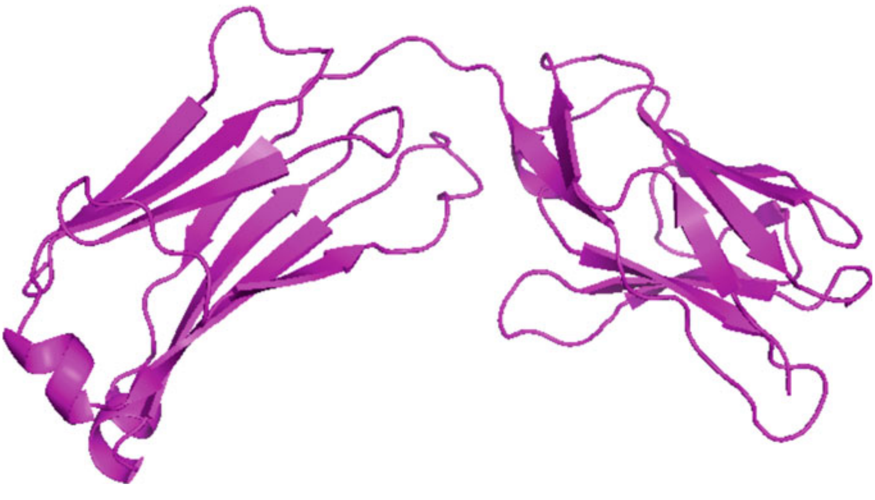
R. Oriama (✉) · T. Burger-Helmchen
University of Strasbourg, Strasbourg, France
e-mail: burger@unistra.fr

R. Oriama · R. Mudida
Institute for Public Policy and Governance (SIPPG), Strathmore University, Nairobi, Kenya
e-mail: rmudida@strathmore.edu

R. Oriama
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Strasbourg, France

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_12

for biosimilar development include transdisciplinarity, harmonization, value chain diversifications, intellectual property, and knowledge diffusion. We describe the landscape of biosimilar development across different economic regions and present case examples of health research development and the place of incremental innovations. We elaborate on how regulation and public innovation policies become part of a multi-level governance system, especially in coordinated market economies. Presented themes for multi-level systems and transitions towards a health bioeconomy through biosimilar development involve different pathways leading to transformation, de-alignment, realignment, technological substitution, or reconfiguration. Dynamic interactions between the landscape, regime, and niche level ecosystems alongside the varieties of innovation contribute to this transformation. Finally, we recognize the role of regulatory harmonization in health preparedness and universal coverage. Biosimilars development has emerged as a strategic niche for unlocking the biophysical characteristics of nature. Learnings from complex multi-level systems can allow firms, governments, and individuals to invest in the biosimilar value chain innovatively.



3D art representation of adalimumab (C6428H9912N1694O1987S46). The crystal structure for adalimumab rendered on Predicted on CHpred and rendered on PyMol.

(Adapted with permission from Shadrack Odikara).

The code of this chapter is 01000010 01101001 01101111 01110011 01101001 01101101 01101001 01101100 01100001 01110010 01110011.

1 Introduction

Biosimilars present an opportunity for incremental innovation in biopharmaceutical development, a component of the health bioeconomy. Incremental innovations relate to the implementation of the continuous improvement paradigm to products and processes, a process that has become the hallmark of biosimilar development, where quality control levels are the most significant determinant of successful development, as any variations in the production process may lead to the formation of entirely different products. Biosimilars are biologics that are "similar" or copies of the original but still performing the same function and are often much cheaper. Their existence spells numerous advantages over their precursor biologics. Biosimilar development has seen tremendous success in coordinated market economies, especially in Europe, vis-à-vis liberal market economies, which are more propice for radical innovations [1]. Emerging economies have made a significant effort in biosimilar development with the characteristic of more mixed-type economies, although distinctions can exist along the continuum.

Considering the level of research and development (R&D) and intellectual property configurations in health pharmaceuticals and biologics, firms that developed products earlier in time, predominantly in the 1970s, have had significant advantages in claiming dominance and increasing the barriers to entry for new firms. Production was limited through registered patents, and any firm that wanted to create the pharmaceutical product had to license in production rights. Since then, with the increasing expiry of the patents, there emerged new opportunities for entrant firms. The first biologicals patented in the early 1990s already have lost patent protection, offering cheaper biosimilars opportunities. This loss of biologics patents, predicted to account for 50% of global pharmaceutical sales, paves the way for biosimilars market growth, which is poised to penetrate global markets influencing innovation networks even in emerging economies [2, 3]. The need for cheaper health solutions echoes global efforts for universal health care as governments emphasize strengthening their health systems.

With the advancement of research, technology, and scientific knowledge, the pharmaceutical industry redirected R&D efforts towards developing novel biologic products, having the first ones authorized in the 1980s. These products included enzymes, protein-based hormones, monoclonal antibodies, blood products, vaccines, genes, and cellular therapies. These treatments made use of sources including humans, microorganisms, or animals. The advent of the genomic era, the New Biology, and advancements in modern technologies coupled with the complex social and economic structures governed by regulatory bodies has accelerated the need for innovative biosimilars to supplement the high costs associated with the biologic counterparts. With patented biopharmaceuticals, the processes become fixed, giving any new manufacturer a limited operating framework. In the new opportunities, we find that the biological processes create variability. These variations can be in the form of new products, substantial changes in the production line used or producing the drug entirely. An example is Amjevita (adalimumab-atto), derived from the reference biologic Humira (adalimumab). It is a monoclonal antibody (mAb) classified as a TNF (tumour necrosis factor) inhibitor used to treat different diseases, mainly rheumatoid arthritis and psoriasis.

The Biosimilar development process takes on a transdisciplinary nature due to the roles the different actors play, from preclinical development to clinical phase development to the markets. These roles include those in molecule development, early product development, testing, manufacturing, trials, and commercialization. These diverse actors are working within this value chain, respect the rules of the global value chain, where considerations for sustainability and wicked problems cannot be neglected. Transdisciplinary work towards engaging directly with production and knowledge outside academia, thereby moving beyond academia's divides. From this perspective, the research approach's core is the societal impact [4], interacting in a transdisciplinary fashion with landscape stakes that influence the multi-level perspective (MLP). For instance, access to universal health care as championed through the sustainable development goals (SGDs) affects health R&D and policy interventions to impact this societal need.

The biopharmaceutical space has been experiencing high attrition rates, where a drug has to go through the preclinical to the four clinical phases to demonstrate safety and efficacy. Studies have reported failure rates of about 30, 60, 30–40%, and the teens in Phase I, Phase II, Phase III, and at the NDA-to-approval stage [5].

There have been health care and regulatory pathways considered as creating barriers, for example, in the production of advanced therapeutics like regenerative medicine (RM) and stratified medicine [6]. These therapies do not have ascertained routes to market and would transform existing pathways if they are successful. The regulatory frameworks, though variable across different economies, made biosimilars development to adapt varied configurations, still with significantly reduced costs. In the European Union (EU), for instance, these have to show pharmacokinetics ADME i.e. absorption, distribution, metabolism, and excretion, hence similarity with the reference biologic to be approved for use. Clinical studies evaluating potential biosimilars aim to confirm efficacy and safety and not necessarily demonstrate clinical benefits. This configuration makes it such that they are designed differently from the clinical studies targeting novel biologics' approval. On the contrary, since these were established by the originator, biosimilars do not require far-reaching studies to evaluate action mechanisms, determine optimal dosing, or demonstrate patient benefit [7]. In some instances, biosimilar approval may be granted using simple data extrapolation without directly evaluating clinical studies' indications (Fig. 1) [7, 8]. This class of drugs demonstrates one of the most expensive cost treatments globally, implying the need for low-cost options. Biosimilars are more affordable, making them attractive and essential to economies, especially when treatments are expensive and so remain not available for many people. The availability is due to the lack of affordability in lower per capita Gross Domestic Product (GDP) levels in emerging countries and higher absolute poverty rates [8].

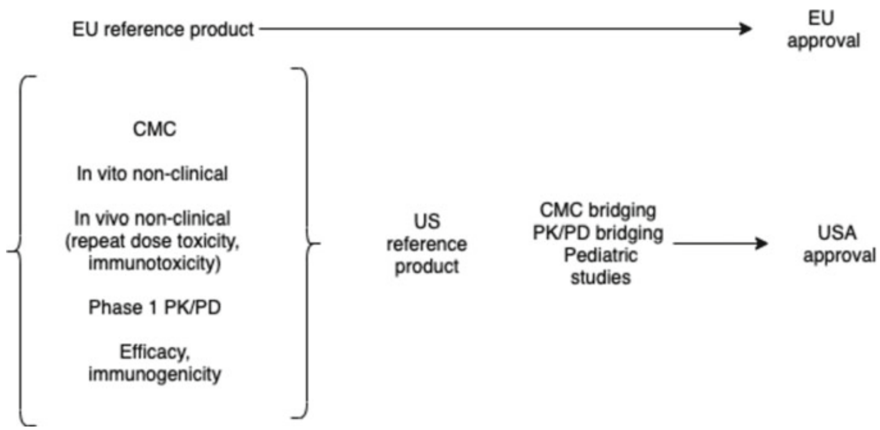


Fig. 1 Stages of biosimilar regulation and approvals, the EU versus the USA [8]

The bioeconomy can be leveraged to provide a competitive advantage for economies and firms involved in the biosimilar value chain. The Global Bioeconomy Summit 2018 defined bioeconomy as "*the production, utilization and conservation of biological resources, including related knowledge, science, technology and innovation to provide information, products, processes and services in all economic sectors aiming towards a sustainable economy"* [9]. The bioeconomy is a project to improve national competitiveness by leveraging valourization and clever use of local biological resources and their knowledge to compete in global markets and a technological project to meet global human-security challenges like food security, climate change, and health [10]. Countries worldwide are developing policies and strategies and aligning them to tap into their biological resources to gain competitiveness. So far, about fifty states have placed the bioeconomy in their political agenda [11]. It means that while determining their growth and development trajectories, more and more economies recognize the roles of biological recourses and their neoliberalizations through the real subsumption of nature as captured in the concept of the knowledge-based bioeconomy or through the formal subsumption of nature [12]. A processual approach by Birch and colleagues considers neoliberalization as "*extensive deconstruction and reconstruction of institutions, often in the name or in the image of 'markets."* Such an approach leads to *"a tangled web of state-regulated oligopolies, profit-orientated enclaves and pseudo markets*," instead of the common understanding of the word referring to "free-markets" that operate with reduced state intervention. In Africa, South Africa has captured health as one of the defining sectors for their bioeconomy [13]. It has stood out as the most innovative country in the continent in health innovation, followed by Kenya. This is a ranking offered by the Global Innovation Index (GII) "*that assesses based on Institutions, human capital and research, market sophistication, business sophistication, knowledge and technology outputs and finally technology outputs"* [14]. The Eastern Africa region is in the process of building its bioeconomy strategy through the "Developing and innovation-driven Bioeconomy Strategy for eastern Africa (BiSEA)" project, set to be completed in 2021.

The Bioeconomy presents an opportunity to solve complex problems. Health challenges as captured in the 3rd United Nations Global Goal include infectious diseases, communicable diseases, health systems, and funding. These challenges can be addressed not only through interdisciplinary focus but also through political strategies. Science diplomacy, anchored on the sovereignty of states, speaks to these challenges as most recently witnessed during the COVID19 global epidemic where nations worked and increased commitment to reinforcing their internal capacity and resilience for combatting the epidemic through science. There have also been efforts in regional and global collaborations. Often, such challenges as seen in such health crises are universal and often necessitate borderless solutions where exchanges between researchers can represent unique linkages where science becomes the vanguard of diplomacy [15]. Globalization has contributed to the increasing spread of the pandemic, hence making global solutions relevant. This kind of support, particularly to basic research, has ripple effects on economies owing to the non-rival and non-excludable nature of knowledge, a common good.

Involving such landscape and regime-level dynamics to biosimilar development processes fosters the substitution of affordable medicine and biomolecules in in-demand markets worldwide. Other outcomes for such efforts include the value chain side streams that result in job creation, technology transfer, and the advance of emerging markets towards the technology frontier. Recent trends in trials and commercialization have seen firms negotiating for partnerships to increase visibility and focus productivity. The business model is moving towards firms that develop products up to the cheaper preclinical phase. After that, they would license their products to larger firms. This model follows the connections that progressively found barbell-shaped ecosystems witnessed in the pharmaceutical domain. Cross-division networks result from a balance by a blend of smaller start-up-like companies on the one side with unique digital and data capabilities. On the other side of the barbell are a couple of large incumbents—these work to drive the commercialization of novel biological technologies [16]. Although accessing biosimilar emerging markets could be a less attractive strategy compared to tapping into an established, developed market of high-margin originators biologic players, winning in these markets can help position biosimilar companies for long-term success. We proceed to describe the landscape of biosimilar development across the different economic regions in the second section. Following this description, we argue in the third section that health research development and incremental innovations, elaborating how regulation and public innovation policies become part of a multi-level governance system, especially in coordinated market economies. Reflecting on best practices from global markets, we demonstrate in the fourth section how the borderlessness of this opportunity helps in solving complex problems common in health research development. The fourth section reflects on lessons for emerging economies and reverses innovation. Finally, three sets of key messages and recommendations are forwarded.

2 Biosimilar Development

The New Biology, coupled with pharmaceutical engineering, has in the last thirty years become the dominant knowledge in the biopharmaceutical industry, building on biotechnology. This knowledge has diffused globally, having scientists proclaim that the biological sciences will define the twenty-first century. The concept of New Biology is understood as originating from discovering the double helix structure of the DNA by Watson and Crick in the 1960s. This new knowledge allowed for gene mapping of entire organisms [17]. Biologists mapped the human genome in the 1990s, a success that promised hope in the understanding of the human template and disease interventions. Further, with technologies such as stem cells and CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-associated protein 9), gene editing is more and more accessible. When coupled with the advent of information technologies, biological innovations of the moment and the future have been projected to cover four areas, according to a McKinsey Report on The Bio Revolution: Biomolecules, Biosystems, Biomachine interfaces, and Biocomputing. Biomolecules map cellular processes and functions by measuring intracellular molecules (e.g. DNA, RNA, proteins); its engineering involves intracellular molecules, such as genome editing for gene therapy for monogenic diseases. Biosystems, on their part, map complex biological organizations, processes, and interactions among cells; it involves the engineering of tissues and organs using stem cell-based technologies and transplantation. Biomachine interfaces help enlighten us on functional and structural aspects of the central nervous system, its engineering involving hybrid systems that provide the nervous systems of living organisms with machines to generate neuroprosthetics for motor control. Finally, biocomputing maps intracellular pathways or cell networks to associate outputs with the condition of computation, its engineering needing data storage, retrieval, and processing about cellular and intracellular measures, for instance, about strands of DNA [18].

A common good for states and individuals, the diffusion and adoption of this knowledge and its technologies have been widespread, with economies seeking to have more of their genetic and biological resources mapped out. Owing to the configurations of international intellectual property rights (IPR) systems and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) of 1995 that has for the first-time linked questions on intellectual property with questions on multilateral trading minimum standards on IPR were set. Much of the pharmaceutical-related patents are concentrated within the Global North, often within the “IP Triad” among North America, Europe, and Japan. In addition to the TRIPS Agreement, the internationalization and globalization of the IPR system through the Paris Accord of 1960 has opened up countries that have ratified these agreements open to collaborations with firms holding active patents. Patent publications allow firms to replicate processes and develop biosimilars against respective reference biologics, representing just about 50% of all biopharmaceutical products globally [2]. This trend has extended across the world, with

Europe leading. The United States (US), owing to the liberal nature of its market economy relative to the coordinated market economy, as will be elaborated later, has had delays in the biosimilars development [19]. For the first time in South Africa, Cipla, an India-based pharmaceutical company, announced in 2018 its ambitions to set up a facility in Durban on biosimilar manufacturing [20].

Expiring patents are creating the opportunity for biosimilar development. Due to the cost-effective nature of their development vis-à-vis biologics, these go a long way in addressing complex problems, especially concerning access to medicines and universal health care. More particularly, there has been an increased need for developing in-house solutions to in-house challenges. The patterns of epidemiology have not followed a universal trend. A transition is epidemiologically defined as *"a time interval, with a beginning, when infectious diseases were predominant, and an end, when non-communicable conditions have finally dominated the cause of death."* However, as exposed in [21], these projections do not hold for all populations, as is the case for African populations. Therefore, with differing health priorities across economies, there indeed is a case for developing in-house health solutions. Providing an example, what we expect in the demographic, epidemiologic, and health transition frameworks is a decline in mortality and fertility rates accompanied by an increase in life expectancy trajectories over the 60 years post-independence period. However, Defo observes a move from a homogenous pattern of populations with uniform mortality and fertility pre-independence period to a heterogeneous pattern post-independence. These are patterns influenced by factors affecting the history and experience of countries in the continent, including conflict and instability, disease, economic factors, and health-care systems.

As will later be elaborated, the interest for such an approach where in-house solutions occur is that even more supported by the borderless characteristic of health care. Health conditions affecting a population are very much likely to affect others despite different environmental and socio-economic situations. Increasingly, emerging economies, especially in Asia, provide an attractive market for establishing biopharmaceutical manufacturing facilities due to the population dynamics, access to a knowledgeable workforce, international IPRs, and cheap labour. Besides, Asia is experiencing rising incomes and declining poverty rates, with many countries achieving middle-income status, important in providing an attractive market. Both the big pharma and mid-sized firms find their way to these markets to curve and new niches. Such collaborations facilitate the reduction of the financial burden on health care by reinforcing linkages to global networks for biosimilar development, manufacturing, and commercialization. Some of the expiring patents indicated by their International Nonproprietary Names (INN) and Reference products address these diseases, as enumerated in Table 1. The INN distinguishes the biosimilar from its reference biologic, with different regulatory bodies having different naming requirements [22].

In response to the increased demand, governments are developing relevant policies and regulations accordingly to support these processes. Biosimilar regulatory environments follow classical drug development life cycles with modifications that allow for assured quality control. Quality is a condition sine qua non to

Table 1 Exclusivity expiration periods of selected biologics (Prepared with data from [7])

		Exclusivity expiration	
INN Name	Reference product	EU	USA
Adalimumab	Humaria	2018	2016
Etanercept	Enbrel	2015	2028
Infliximab	Remicade	2015	2018
Bevacizumab	Avastin	2022	2019
Trastuzumab	Herceptin	2014	2019
Pegfilgrastim	Neulasta	2017	2015
Ranibizumab	Lucentis	2022	2020

biosimilar development. In this regard, the general regulatory paradigm involves demonstrating similarity between the reference product and the potential biosimilar, judged against safety, purity, and efficacy. Clinically, there should not be any significant differences.

In the EU, the European Medicines Authority (EMA) is the main regulatory body that harmonizes the regulatory requirements of member countries. In vivo toxicology, studies are not a routine requirement, but often in vitro, evaluations are performed to ascertain the structure–function relationships. There are policy guidelines provided in the 2001/83/EC Directive, Article 10.4. In the US, the American Food and Drug Administration (FDA) is the regulatory body, where it requires that in vivo, comparative toxicity studies be conducted. Their review process stipulates the biosimilar structure alignment with the reference biologic, with multiple levels of supervision and oversight. The policy guidelines for these are provided in the Biologics Price Competition and Innovation Act (BPCI Act) of 2009.

The EU distinguished itself as the first body to have explicit biosimilar policies, pioneering biosimilars' regulatory requirements in 2005 [23], and authorizing biosimilars for the market. Its robust regulatory processing has been attained in part due to the extensive experience with licensed biosimilars. In October 2014, the Committee for Human Medicinal Products revised its guidelines. The current trend is of biosimilars receiving approval in Europe before in the US. The first US biosimilar Zarxio (filgrastim) stimulated biosimilar development in the country, with the FDA releasing final biosimilar guidelines in April 2015.

The Asian biopharmaceutical industry developed quickly, as much as it started much later than the West. There are currently more biosimilar products now in development and use across the Asia Pacific region than elsewhere. However, most of these current products fail to meet the rigorous global criteria of biosimilars, hence could not be distributed globally. This shortcoming earned them the name "follow-on biologics or intended-to-copy." The regulatory body for the technical evaluation of biosimilars, the Centre for Drug Evaluation (CDE), does not characterize compatibility or an opinion on substitution in its guidelines. Instead, it suggests that what is necessary is a product-specific extrapolation of indications. As for India, in 2007, Reditux, replicated from Mabthera, was the world's first biosimilar monoclonal antibody. There are now more than fifty biosimilars in the market in India and beyond.

The biologics market often falls into four primary regions: Europe, America, the Middle East & Africa, and the Asia Pacific, the latter being the fastest growing, with a 5.5% Compounded Annual Growth Rate. The United States has the largest market. Africa still is the least market for biological drugs [24]. A global study by McKinsey states that an estimated $5 billion–$8 billion of biosimilar sales would emanate from emerging markets, with the total sales doubling from 2014 to 2020. The biosimilar regulatory framework landscape is changing in the African continent at different stages of maturity. So far, South Africa is the only country that has established a clear regulatory framework for biosimilar approval. Soon, one can expect related legislation introduced in Algeria, Morocco, and Tunisia, adopt the recommendations of the EMA, WHO, and the FDA concerning biosimilarity, quality, efficacy, and safety standards. Kenya, through the Health Act 2017, is committed to forming a body to regulate the practice of alternative medicine.

3 Health Research Development and the Place of Incremental Innovations

Biosimilars differ from generics in that in the patented medicines; the processes had been fixed, giving any new manufacturer a limited operating framework. In the new opportunities presented by biosimilar development, the biological processes create variability and room for innovation, allowing for incremental innovation. Incremental innovation differs from radical innovations where revolutionary products or processes come to exist. Advances in health care occur parallel to technological progress, be it an advanced form of imaging cells or a faster diagnosis method. Technology continues to drive improvements in health care, the revolution being catapulted by biotechnology advancement and advances in information technology, where the latter has mainly provided the impetus in incremental process innovation. Products innovation generally refers to developing a new product, improving product performance, or a new feature in a product. Process innovation, on the other hand, generally refers to novelty in procedures, including associated technologies, skills, or facilities. The ability to link doctors and hospitals, rural patients with healthcare professionals, or link researchers to one another transforms health care.

Regulations and public innovation policies are influential, especially to the landscape and regime levels of multi-level governance systems. The socio-technical transition theory is a heuristic framework that comprises, among others, the MLP framework, which is concerned with successful transformative processes, and varieties of innovation. The MLP framework has been expanded to describe transition pathways, spanning from transformation to de-alignment and realignment, to technological substitution, to reconfiguration [25].

Technologies, institutions, and social and economic subsystems can come and perform co-evolutionary development under this MLP framework. There are three levels to understand transitions: the landscape level (macro), the regimes level (meso), and the niches level (micro). The landscape-level encompasses the macro-economic and macro-political development that shape socio-economic transitions, for instance, the political context of an economy. The regime level entails institutions and structures that reinforce the socio-technical systems, whether these be individuals, firms, states, or social norms that influence, for instance, the IPR and competition laws of a complex system. The term "institutions" in this case is understood from a neo-Schumpeterian lens where institutions function to dictate firm behaviour and the selection environment, leading towards technological change [26]. The niche level on its part is attributed to the development of radical innovations from experimentation. This is a less coordinated level, unlike the previous two, with less regulatory influences. Different systems will adapt different transformation pathways following the MLP.

The transformation pathway of MLP is characterized by moderate changes in the socio-technical landscape, with niche innovations not being fully formed. In this case, the current regime will adapt to the niches' changes; however, no substantial transition occurs. In biosimilar development, regulations and policies are sometimes in harmony with dominant global trends, but real interest exists to support niche and regime-level innovations.

The de-alignment and realignment pathway occurs when the landscape change is significant and abrupt but are niche innovations not fully formed. Multiple niche innovations co-exist and compete until only one innovation remains, replacing the old technologies. If the biosimilar domain were to follow this pathway, current health care and regulatory pathways would require accommodating the needs of advanced therapies and biosimilars. The traditional methods would be completely adaptive to allow for replacement with new therapies, which may not even yet have been discovered.

In technological substitutions, the landscape creates significant pressure on the regime when fully formed niche innovations are available. These innovations immediately replace the old technologies. In this pathway, complete adoption of biosimilar development would occur. However, setting up infrastructure and a knowledge-based labour force beforehand is necessary due to the high-tech nature of the domain. This pathway has been deemed the least probable in high-tech domains.

Finally, in the reconfiguration pathway, symbiotic innovations develop in niches. These innovations are initially adopted to address local challenges in the regime. Consequently, these innovations set off further modifications to the system. This transformation pathway has been the case for some of the biosimilar products in Asia, where due to established non-conformity to international standards, these products have been dubbed "copy-biosimilars" and consequently have reduced marketability. Subsequently, resulting landscape configurations consider in-house challenges of production, manufacturing, and commercialization and make room for even more competitive niche-level innovations.

Intellectual Property provides legal protection for innovators. In the biopharmaceutical domain, the product receives an INN and a brand or trade name selected by the innovative pharmaceutical company. Patent protection allows the patent holder to market the new substance for a given duration [22]. As patents can be issued to individuals or businesses, there are differences in the set of rules used in granting patents.

Coordinated market economies coupled with industry interrelatedness have been linked to technological development success, especially for biosimilars. As argued in [27], quality control is a condition *sine qua non* in biosimilar development. Small variations in production processes can result in entirely different products with different characteristics. Their advantages notwithstanding ever-limiting patents in pharmaceutical industries have for many years in varying degrees restricted the innovation processes and the subsequent development of new frontier cheap and affordable medicine. Even though the room for innovation and development has expanded, the depth of knowledge on biological diversities in the advent of the genomics era has widely expanded the arena for biosimilars and generics developments. The vast array of genomics data generated due to the incorporation and use of modern-day genome sequencing techniques, which has improved output over the years in terms of cost, quantity, and quality, gives way for even [28]. Economy-specific requirements impact the ease of adoption of technologies and processes. While liberal market economies are more expedient for radical innovations and coordinated market economies for incremental innovations, the latter allows for de-alignment and realignment pathways. While niche innovations are in production, the landscape and regime levels align themselves to suit the needs of these niche innovations, allowing for organic growth of these technologies and their adoptability in markets. Such cohesion is what has mostly the case in the EU and can be attributable to witnessed success.

4 Best Practices from Key Global Markets

Arguably, the shift from radical innovation in biologics to the incremental innovation of biosimilars witnessed in the EU has transformed the productivity of the biotechnology and biopharmaceuticals domains. This shift has resulted in a significant evolution of firm strategies for competitivity and meeting health demand conditions. As of 2014, Germany has taken up to 33.1% of biosimilar market share in the EU. Globally, biologics and associated products account for just about 50% of all biopharmaceutical products. This transformative growth rate is mainly attributable to the policy and governance structures in these economies that have helped adopt acceptance and approval of generics and biosimilar pharmaceutical drugs [1].

The newfound ability of entities to collaborate and compete globally in a borderless or inter-jurisdictional environment drives globalization in this age. As communities and populations are involved more and more into discourses no longer solely led by governments and large corporates, but through a citizen science approach, stakes are being raised even higher. In health care, borderlessness has further been accentuated in phenomena such as the migration of climate-sensitive infectious diseases (CSIs) towards the North, with climate change affecting both the epidemiology and geography of these diseases, which are of relevance to both humans and animals. Secondly, with increased migration, infectious diseases are spreading even faster, as evidenced by recent COVID and Ebola infections. To cite the third situation of relevance, as argued by Prof Jefferey Sachs in the Age of Sustainable Development, income inequalities across the world have been attributed to the increases are responsible for social and environmental inequalities associated with the planetary boundaries witnessed today [29].

Biosimilar value chains provide solutions to the immediate challenge of access to affordable medicines, thereby addressing questions on universal health care for all, especially for vulnerable populations. From learnings, coordinated market systems are more useful for such incremental innovations by proving frameworks through policy and regulation that enhance the level of quality needed to ensure that they find a quicker way to the markets. This market coordination coupled with the national pricing systems provides medicines for a wider population.

5 Lessons for Emerging Economies

International innovation has a long history where emerging and developed economies can learn from each other [30]. Vernon [30] describes the product life cycle model from a global perspective to explain the trade between the United States and Europe after the Second World War. Specifically, his purpose is to understand why and how US companies develop their products in international markets (mainly in developing countries) through foreign direct investment. When Vernon published his theory in the 1960s, the United States was the world's leading economy and the primary product innovation source. New products were initially developed for the massive US domestic market and then introduced to other countries.

Vernon identified four stages for the company to go international: Introduction, growth, maturity, and decline. In the introduction phase, production is basic, and the target market is domestic. At this level, there is no need to go abroad for sales or production. The company benefits from a specific monopolistic position because it sells brand new products with high growth potential and limited competition. The focus here is to penetrate the domestic market. In the second stage (growth), there are products primarily accepted by consumers. Sales have grown not only in the local market but also abroad. The products are exported to foreign countries, thereby gaining market share. In the mature stage, the company began to face price competition from local competitors, and local competitors entered the market due to strong demand. It is no longer possible to export products from the country of origin. This situation requires foreign direct investment in production facilities to benefit from lower local production costs.

In the 1970s, Vernon realized that his model would become outdated. International markets are seeing emerging economies such as Japan and Europe. In this case, most new technologies come from countries outside the United States. In 1974, Vernon introduced a revised version of its product lifecycle model, which considered oligopolistic competition. More recently, many scholars pinpointed that innovations are improved by or even introduced by emerging countries and that the diffusion of innovation may sometimes be "reversed" [31, 32]. Scholars showed that even in the specific domain of health science, emerging economies could provide lessons to the more advanced countries, disrupting even more Vernon's old diffusion theory [33, 34].

Several adoption lessons can be drawn from an MLP to biosimilars development, pathways towards incremental innovation in the health bioeconomy. To begin with, interest in biosimilars in emerging economies does exist, backed up by the need to create in-house solutions and a ready workforce to participate in the development, manufacturing, and commercialization of these products. These conditions explain the presence of "copy-biosimilars," which are predominantly produced in Asia being marketed in Africa, due to an existing weak link between policy, regulation, and practice. Having established the interest and need comes how landscape, regime, and niche systems can be organized to favour the production of high-quality biosimilars and their integration into the health systems. Firstly, as evidenced in generic medicine production, Asia, notably India, led the way. Through step-by-step infrastructure development, they have built systems that position the economy as a leader in the domain. The flying geese analogy can be employed where an economy in a region inspires other related economies towards socio-economic development. This phenomenon in Japan led to advancement towards the technology frontier in the last 30 years. Soon after, South Korea, China, Singapore have risen as growing leaders in knowledge-based technologies.

Initial steps towards developing infrastructure can inspire regime and landscape-level systems to adopt and set up policies and regulations in support, following the reconfiguration pathway, where symbiotic innovations occur in niches. Such configurations are initially adopted to address local challenges in the regime, setting off further regime modifications. This pathway is more suitable for systems like in China and India, where there have been significant efforts invested in biosimilars. Still, regime-level institutions have not adequately caught up to allow for international standards.

The African Union is establishing the African Medicines Agency (AMA), which will primarily adopt guidelines from the World Health Organisation (WHO) and the EU since May 2018. The African biosimilars market is in its infancy. Still, the potential is enormous. This strategy may either follow the transformation pathway of MLP where there are moderate changes in the socio-technical landscape with niche innovations are not fully formed. It may also follow the de-alignment/realignment pathway when the landscape change is significant and abrupt, with niche innovations not fully formed. The lack of policy implementation is an epidemic that scourges many governments and economies, and the latter pathway may add to this narrative. Owing to the increasing demand for cheaper pharmaceutical options and the pharmaceutical networks in the continent, countries can leverage these global partnerships to continually develop infrastructure for manufacturing, at the same time developing needed regulations and policies.

6 Conclusion

The health sector presents opportunities for innovation and is poised to transform how business is done globally. Some recommendations can be forwarded to advance robust multi-level systems for biosimilars development. These recommendations are tied with health bioeconomy principles stemming from the conviction that these principles, through the real subsumption of nature, support the sustainable exploitation, use, management, and conservation of natural resources providing the interface between modern and indigenous knowledge and advances in technology for innovations.

Harmonization of regulations provides added support to health preparedness and universal health coverage. Weak links caused by a lack of policies or hesitant regimes can hinder the development of advanced therapies. At the same time, precocious development of these policies without adequate development of infrastructure and interest at the niche level may not provide the transformative capacities intended. Current health care and regulatory pathways need to recognize the necessity for advanced therapies such as biosimilars, regenerative medicine, and stratified medicines and develop pathways that accommodate them. The need for these therapies is more often than not demand-driven, and relooking their pathways may contribute to better health preparedness.

The knowledge-based bioeconomy works on the neoliberalization of knowledge through strategies, frameworks, and actions which intensify real subsumption of nature through state-led market exchanges. As such, building capacity for a knowledge-based bioeconomy can be led by both demand and supply side interventions. The recognition of the impetus of this kind of a bioeconomy supports technological innovation, leveraging on the renewable, biophysical characteristics of nature. This transformation, described as a techno-economic fix, allows for progressive growth and sustainable capital growth. On developing a knowledge-based bioeconomy, evolutionary economists project the emergence of new sectors by putting in place bioeconomic technologies. At the same time, the development dynamics of some of the then-existing industries will get fresh momentum i.e. the dedicated innovation systems. Such transformations happen in the case of biosimilars or even the circular economy industries, algae and plant-based bioplastics, bio-based textiles, plant-based proteins, among others. A shift to qualitative development from quantitative growth is what will make a difference in a bioeconomy transformation.

Building capacity for innovation by firms is not only limited to biosimilar development but can extend to manufacturing and commercialization. Any such efforts have eventually been found to prepare niche-level innovations by systemically building infrastructure that allows for scaling up and diversification of value chains' involvement. Borderlessness, concerning the sovereignty of nations and the reach of complex global problems, has allowed for collaborations to be negotiated, creating technology platforms distributed with less and less concern for geographical limitations. IPR systems, albeit their restrictions, allow for R&D cooperation and licensing agreements to be made that favour specialization. Firms, governments, and individuals need not hesitate to get involved in the biosimilar value chain and its side streams. Learning from the complexity associated with these multi-level systems provides an excellent avenue to build confidence in terms of forging stronger global collaborations, improving infrastructure, establishing harmonized policies and regulations, and implementing these.

Core Messages

- The biosimilar development process takes on a transdisciplinary nature due to the roles the different actors working within this value chain play with respect to the rules of the global value chain, where considerations for sustainability and wicked problems cannot be neglected.
- Current health care and regulatory pathways need to recognize the necessity for advanced therapies such as biosimilars, regenerative medicine, and stratified medicines and develop pathways that accommodate them.
- Building capacity for innovation by firms prepares niche-level innovations by systemically building infrastructure that allows for scaling up and diversification of value chains' involvement.
- With the increasing expiry of biologics' patents, new opportunities for entrant firms predicted to account for 50% of global pharmaceutical sales would pave the way for biosimilars market growth and provide cheaper health solutions.
- A common good for states and individuals, the diffusion and adoption of knowledge on biosimilar development and its technologies have been widespread, with economies seeking to have more of their genetic and biological resources mapped out and exploited for competitivity.

References

1. Pyka A, Buchmann T, Vermeulen B (2017) Biosimilars in Germany: the emergence of a new industry in the light of the varieties of capitalism approach. Technol Anal Strateg Manage 29:276–289. https://doi.org/10.1080/09537325.2016.1262022
2. GaBI Journal Editor (2019) Patent expiry dates for biologicals: 2018 update. Generics Biosimilars Initiative J 8:24–31. https://doi.org/10.5639/gabij.2019.0801.003
3. EvaluatePharma (2014) World preview 2014, Outlook to 2020

4. Toomey AH, Markusson N, Adams E, Brockett B (2015) Inter-and Trans-disciplinary research: a critical perspective
5. Lowe D (2018) A new look at clinical success rates. In: The Pipeline. Science Translational Medicine
6. Mittra J (2016) Regulation, policy, and governance of advanced therapies. In: The new health bioeconomy. Palgrave Macmillan, US, pp 121–156
7. Declerck P, Danesi R, Petersel D, Jacobs I (2017) The language of biosimilars: clarification, definitions, and regulatory aspects. Drugs 77:671–677
8. Dolan C (2018) Opportunities and challenges in biosimilar uptake in oncology. Am J Manag Care 24:S237–S243
9. International Advisory Council of the Global Bioeconomy Summit 2018—IAC-GBC (2018) Global bioeconomy summit conference report. In: Innovation in the global bioeconomy for sustainable and inclusive transformation and wellbeing. 108
10. Mittra J, Zoukas G (2019) Unpacking the concept of bioeconomy. In: Problems of definition, measurement, and the attribution of 'value'
11. Global Bioeconomy Summit (2018) Global Bioeconomy summit communiqué 2018. In: Innovation in the global bioeconomy for sustainable and inclusive transformation and wellbeing, pp 1–24
12. Birch K, Levidow L, Papaioannou T (2010) Sustainable capital? The neoliberalization of nature and Knowledge in the European "knowledge-based bio-economy." Sustainability 2:2898–2918. https://doi.org/10.3390/su2092898
13. Pugatch Consilium (2014) Building the bioeconomy. Examine national biotechnology industry development strategies
14. World Intellectual Property Organization (2019) Global Innovation Index 2019: summary
15. Ruffini P-B, Ruffini P-B (2017) Science in the vanguard of diplomacy. In: Science and diplomacy. Springer International Publishing, pp 85–104
16. McKinsey Global Institute (2020) Driving a new generation of biological technologies| McKinsey. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-the-bio-revolution-could-transform-the-competitive-landscape?cid=other-eml-alt-mcq-mck&hlkid=d107120850ce4db9ba56e79be18506dc&hctky=2845898&hdpid=fd6b86ea-3273-47a2-a275-428c882. Accessed 20 Jun 2020
17. Mittra J (2015) The new health bioeconomy R&D policy and innovation for the twenty-first century
18. McKinsey Global Institute (2020) The bio revolution, innovations transforming economies, societies, and our lives
19. Rémuzat C, Dorey J, Cristeau O et al (2017) Key drivers for market penetration of biosimilars in Europe. J Market Access Health Policy 5:1272308. https://doi.org/10.1080/20016689.2016.1272308
20. Pharmaceutical Technology (2018) Cipla biosimilars manufacturing facility, Durban—pharmaceutical technology. https://www.pharmaceutical-technology.com/projects/cipla-biosimilars-manufacturing-facility-durban/. Accessed 22 Jun 2020
21. Defo BK (2014) Demographic, epidemiological, and health transitions: are they relevant to population health patterns in Africa? Glob Health Action 7. https://doi.org/10.3402/gha.v7.22443
22. Hein L, Mohr K, Bieger D (2005) Color atlas of pharmacology. In: Lullmann H (ed) 3rd edn. Thieme
23. CHMP (2014) Committee for medicinal products for human use (CHMP) guideline on similar biological medicinal products
24. Pategou J (2020) Africa's biosimilar landscape outlook current challenges
25. Vandermeulen V, Van der Steen M, Stevens CV, Van Huylenbroeck G (2012) Industry expectations regarding the transition toward a biobased economy. Biofuels, Bioprod Biorefin 6:453–464. https://doi.org/10.1002/bbb.1333

26. Loiseau E, Saikku L, Antikainen R et al (2019) Technology and institutions in Neo-Schumpeterian and original institutional thinking. Res Policy 53:1689–1699
27. Arbeitsgemeinschaft proBiosimilars (2017) Handbuch Biosimilars
28. Ahen F, Salo-Ahen OMH (2018) Governing pharmaceutical innovations in Africa: inclusive models for accelerating access to quality medicines. Cogent Med 5. https://doi.org/10.1080/2331205x.2018.1500196
29. Sachs JD, Ban K (2016) The age of sustainable development
30. Vernon R (1966) International investment and international trade in the product life cycle. Quart J Econ 80:190–207
31. Hussler C, Burger Helmchen T (2019) Opening the reverse innovation black box to pinpoint its drivers and barriers in Western MNCs. Euro J Int Manage 1:1. https://doi.org/10.1504/ejim.2019.10016818
32. Govindarajan V, Ramamurti R (2011) Reverse innovation, emerging markets, and global strategy. Glob Strateg J. https://doi.org/10.1002/gsj.23
33. Burger-Helmchen T (2019) Reverse innovation in health care. J Innov Econ. https://doi.org/10.3917/jie.029.0217
34. Govindarajan V, Ramamurti R (2018) Reverse innovation in health care: how to make value-based delivery work

Ruth Oriama is an economics and innovation researcher within the health and biotechnology domain, her doctoral thesis focusing on "Innovation-triggered transformation towards a knowledge-based health bioeconomy in Kenya." Her interest is in developing programmes to make a significant difference in building competitiveness and improving human development. Before her current Ph.D. position at Université de Strasbourg and Strathmore University (Kenya), she completed a Master's Programme in Innovation and Research Development at Université Paris Sud (France) and a B.Sc. Biochemistry from the University of Nairobi (Kenya).

Thierry Burger-Helmchen is a professor in Management Science. He is a researcher at BETA-UMR 7522 CNRS. He is the author of more than 40 articles in peer-reviewed journals, and he published several books (textbooks and research books) on economics and management. His research topics are innovation and creativity management (more information at http://www.burger-helmchen.com/).

13 A 2020s Marketing Taxonomy for Augmented Reality Customer Experience

Emad AbouElgheit

One of the areas where consumers adopt something new is when it makes something drastically easier. When I think about augmented reality I think about instances like finding your way somewhere, finding your friends in a stadium, or going to a conference and looking around and knowing who everyone is because their LinkedIn profile is hanging over their heads.

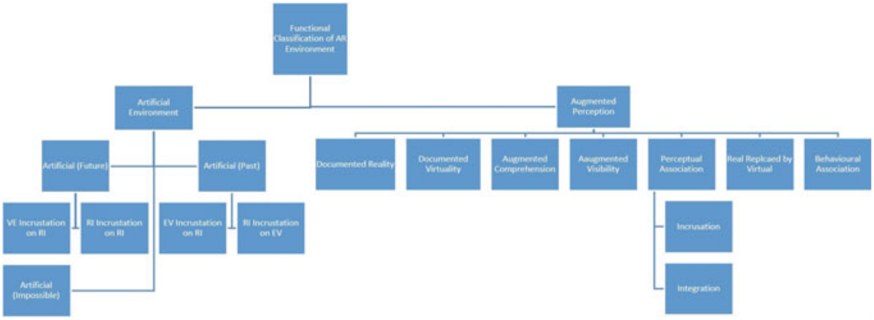
Belsky (2018)

Summary

Metaverse, Virtual, mixed, and augmented reality (AR) applications are already inspiring marketers to engage customers in new types of experiences. Technology-based AR and marketing-based online customer experience literature have been widely disconnected. Plenty of taxonomies for AR applications exist mainly from a technology standpoint. This chapter proposes a theoretical novel taxonomy for AR applications from a marketing and customer experience angle suitable for the 2020s. The author summarizes earlier relevant studies through a comprehensive literature review. The author proposes an AR taxonomy based on the level of immersion in the AR environment (spatial presence) and the type of buying situations. The author suggests four main types of AR experience with marketing strategies and managerial implications for each.

E. AbouElgheit (✉)
Ancell School of Business, Western Connecticut State University, Danbury, CT, USA
e-mail: abouelgheite@wcsu.edu

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_13



Functional classification of AR environment.

(Adapted from [13], https://blog.hubspot.com/marketing/augmented-reality-examples, Accessed: 9 July 2020).

The code of this chapter is 01000101 01111000 01110100 01100101 01101110 01100100 01100101 01100100 00100000 01110010 01100101 01100001 01101100 01101001 01110100 01111001.

1 Introduction

Augmented Reality (AR) will disrupt our daily lives simply because it will make things much easier. AR is a global market expected to reach \$18 billion by 2023, with a massive 1000 percent growth rate compared to 2018 (\$1.8 billion). 67% of media planners and buyers consider using AR and/or VR in their digital marketing strategies [1]. AR users are expected to reach 1 billion in 2020 and advertising expenditure to reach \$2.6 billion in 2023 [2].

AR is the evolution of Virtual Reality (VR), which started as early as the 1950s and gained momentum in the 1990s [3]. While the latter immerses users in an all-virtual three-dimensional (3D) environment, AR integrates the virtual (man-designed) graphics into the real-life setting in real-time. Examples include a virtual callout that pops with information about a historical building that a tourist sees in reality, a virtual map of the nearest car dealers and prices of a car that one sees in the street, real-time face makeup testing, and video chatting filters. Moreover, there is a disruptive and controversial game introduced in 2016, Pokémon Go.

To marketers, AR is an untapped game-changer. It offers new opportunities to reach and engage customers and create different brand experiences. AR enables marketers to meaningfully link the digital (online) offering and resources to the physical point of sale [4]. Brands such as Pepsi, Sephora, Timberland, and Home Depot have already leveraged AR potentials in their marketing strategies since 2012. These include testing home decoration colors and furniture, virtual fitting rooms, smart glasses, immersive outdoor ads, stadium tickets selection, and makeup and accessories testing [5]. However, marketers should not assume that all VR marketing applications influence consumers in the same way. Some VR applications can even negatively impact favorable consumer actions [6]. Also, technology implementation cost and consumers' acceptance still hinder VR mainstreaming [7].

The earliest mature AR applications in marketing started in the late 2000s, synchronized with the iPhone launch in 20,007, by pioneers such as BMW and Enquire magazine. AR applications in marketing have rapidly scaled in the late 2010s, again synchronized with the advancement of smartphone camera capabilities, Snap Chat lenses' popularity among teens, and the evolution of head-mounted displays [2]. By 2020, AR applications in marketing were numerous in retail, gaming, tourism, real estate, and various industries. While marketers mainly aim to improve customers' brand experiences via AR, marketing applications have been more driven and classified by technology (e.g., image recognition and space markers) than improving customer experience [8]. Accordingly, there is a crucial need to propose a new customer, experience-based taxonomy for AR applications. Marketing scholars and practitioners are still struggling to identify precisely how AR can create a different customer experience fundamentally different from traditional marketing [9]. Exploring AR content impacts (i.e., mixed real and virtual layers) in marketing remains under investigation. There is a need for studies that examine the phenomena as a total customer experience, not a customer response [10].

In this chapter, the author proposes a new customer, experience-based taxonomy for AR applications in marketing. The new taxonomy helps marketing practitioners and scholars to utilize the right mix of AR technologies in their marketing and branding strategies. The author starts by analyzing earlier literature on AR and customer experience taxonomies. The author then explores earlier taxonomies of AR applications in marketing as customer experiences rather than kinds of functions or technologies. The author concludes with implications and recommendations for marketers to utilize AR applications to attract and retain customers after understanding the different customer experiences those applications create. Also, the author identifies areas for future research in the field of AR in marketing.

2 Augmented Reality Taxonomies

AR taxonomy literature has been based on the type of technology or the technical function of an AR application. Badkoube et al. [3] identified four main types of AR applications based on the medium or device that activates the experience (i.e., camera, AR-enabled mobile applications, AR in-feed such as a virtual flying banner in a video game, and head-mounted displays). Edwards-Stewart et al. [8] classify AR under two main types: triggered augmentation and view-based augmentation. On the one hand, triggered augmentation happens when the VR application starts working when it identifies a space marker, an object, or a GPS location. On the other hand, view-based augmentation happens when users view an augmented scene and interact with objects in it.

Other literature approached AR taxonomy from more psychological perspectives. As illustrated in Graphical Abstract, Hugues et al. [11] argue that there are two main types of AR experiences based on the ability to go beyond time and place. The first type of AR mainly deals with what can be seen or done in real space and present time (e.g., unveiling an underground station, chatting with a face filter, or real-time object recognition). The second type goes beyond space and time limitations. For example, completing, virtually, the missing nose of the Sphinx statue in Egypt to imagine what it looked like in the past or seeing a futuristic Paris traffic with flying cars, virtually, while looking from the window of a Manhattan office. In this taxonomy, marketers perceive AR strategies as either an enriched present and opportune moment (e.g., revealing the interior of a restaurant they are heading toward) or an imaginative need or want recognition moment (e.g., seeing how that restaurant looked 50 years ago).

Ramírez-Fernández et al. [12] propose a VR taxonomy based on three main factors that affect users' experience (i.e., realism, interaction, and intensity). In this taxonomy, users' experience can be of different pairs: i, real vs. unreal (e.g., the virtual objects are easily distinguishable from real objects); ii, interactive or passive (e.g., many events happen as a result of user actions or behavior); and iii, light or intense (e.g., to what level of details or aggressiveness the virtual objects react in the AR environment).

3 Online Customer Experience Taxonomies

There is a gap in the literature on classifying the types of customer experience in AR. However, plenty of studies examined the types of experiences in electronic commerce. Earlier literature creates taxonomies for the types of online behavior (user experience) and the antecedents or motives that trigger that behavior. Earlier taxonomies are either based on time stages of experience or the types of thoughts and feelings users have during the online experience. Earlier studies leave a gap in examining the impacts of other important variables in AR and online experience in general. These include the type of device, situational factors, and user location.

Rose et al. [13] create a taxonomy where the types of experiences are either cognitive or affective. AbouElgheit [14] proposes a more comprehensive taxonomy where trust and risk antecedents lead to purchase intention rooted in the domain of cognition, affect, experience, and personality. Eigenraam et al. [9] contend that consumers develop five main types of digital engagement. Those are fun, learning, giving feedback to a brand, talk about a brand, and working for a brand.

Pham and Ahammad [15] approach customer online experience from a time stage perspective. They identify pre-purchase, purchase, and post-purchase stages of customer experience in electronic commerce. Pre-purchase experience includes the availability of product information, website appearance, and customization. Purchase experience includes easy checkout and payment options security. Post-purchase experience includes order fulfillment, the responsiveness of customer service, and ease of return.

Fig. 1 User Micro-Moments. Adapted from Google, 2016, https://www.thinkwithgoogle.com/marketing-resources/micro-moments/micro-moments-understand-new-consumer-behavior/

In 2016, Google introduced the concept of "buying micro-moments" (Fig. 1). The term refers to that seconds long when users, mainly mobile, are highly acquisitional shoppers who look for particular information to conduct an urgent task on the go. Google classified those micro-moments into four main categories. "I-want-to-know" moments happen when users need information about a brand, product, or service without an intention to buy (e.g., exploring new restaurants in town for the coming weekend). "I-want-to-go" moments are highly location-based, where users search for a place nearby to buy something or navigate to a location (e.g., explore driving directions for a business meeting). "I-want-to-do" moments happen when users need to get information to help them complete a task (e.g., searching a YouTube video on how to make a table). Finally, "I-want-to-buy" moments happen when users have an immediate intention and readiness to buy something and are looking for information on what to buy or where to get it from" [16].

4 Latest AR Technologies

It is critical to shedding light on the latest AR technologies available and under research and development. Understanding the direction of innovation and the capabilities helps marketers and researchers to have a proactive approach to utilize and perhaps guide that technology into marketing and product development strategies.

Technology giants are becoming more prepared to mainstream AR. In 2017, Apple and Google introduced ARKit and ARCore, respectively. Those are software development kits (SDKs) that enable mobile application developers to create smarter and faster AR-enabled applications [17]. A simple example for non-tech users is that now most smartphones can read QR codes, recognize a check for deposit, and add layers to navigation without a need for a third-party app, as was the case in the 2010s. As of 2020, artificial intelligence (AI) is integrated into AR applications to provide customers personalized recommendations and customer service while in physical retail stores. Major internet browsers such as Chrome and Firefox are already planning on integrating AR into the mainstream web browsing experience [17].

Fig. 2 Hologram use in product display and advertising. Adapted from TrendHunter.com, 2020, https://www.trendhunter.com/trends/hypervsn-by-kinomo

As of 2020, technology companies plan to extend AR's immersive experience to include other sensory dimensions such as touching and smelling. Microsoft researchers have been working on air vortex rings technology, which simulates physical touch by using air pressure from a ring toward users' hands. RemniScent is a company that manufactures neck wearable scent-releasing devices as part of the AR experience [15]. While such applications already exist, Bayern (2019) expects that other senses such as taste will be present as well as all other senses by 2030.

As seen in Fig. 2, brands such as Pepsi, Nike, and Toyota already use hologram technology in marketing and brand activation. Holograms create virtual objects in physical space without the need to wear special glasses or headsets. The technology is still expensive and needs a complex physical system setup [17].

5 Customer Experience in Augmented Reality: Existing Taxonomies

Table 1 illustrates a summary of earlier literature that examine AR applications' impact on favorable customer experience variables and attempt to classify the types of customers' AR experiences.

Table 1 Literature on augmented reality taxonomy in customer experience

Approach	Variables	Author(s)
Impact and correlational	Satisfaction, enjoyment, purchase, attitudes	Javornik [21], Yaoyuneyong et al. [20], Abrar [1], Alimamy et al. [2], Rauschnabel et al. [18], Tsai [25], Greenhalgh [30]
Self-perception	Perceived body image, AR adoption	Yim and Park [23]
Immmersion and belief	User, brand, other users, content, spatial presence	Scholz and Smith, [24]. de Ruyter et al. [26]
Technology involvement	Content, affordability, technology	(Flavián, Ibáñez-Sánchez and Orús [28]
Public and private use	life efficiency, enjoyment, sensual, social, and self-expression	Rauschnabel et al. [18]

5.1 AR Impact on Customer Experience

In general, earlier literature defends the notion that AR has a positive impact on customers' positive word of mouth and brand appeal [18], reduces the perceived risk [19], impacts purchase intention [12], and improves customer satisfaction [13]. Yaoyuneyong et al. [20] compare consumers' perception of hypermedia ads (i.e., users see digital objects embedded on a print ad) compared to traditional print ads. They contend that consumers perceived higher informativeness, novelty, and effectiveness when exposed to AR-based ads. From a more specific approach, Rauschnabel et al. [18] argue that consumers' inspiration is a mediating factor between the benefits they perceive from AR and the changes in brand attitudes. Javornik [21] argues one of the limitations of interactivity in AR applications is mainly machine or content-focused and less human-focused, such as traditional social media.

5.2 AR and Customer's Self-Perception

In retail electronic commerce, AR as virtual fitting rooms is increasingly popular, especially during the 2020 COVID-19 pandemic [22]. However, reality can lead to different and complex human perceptions. Using AR applications in electronic commerce is not necessarily a way to improve customer experience and increase sales. Yim and Park [23] examine customers' perceived body image on their intention to adopt AR-based clothing try-on technology (i.e., see how an outfit on a website would look like on a virtual 3D image of their own body) compared to seeing a traditional website photo of that outfit. They found that users who have an unfavorable body image are more likely to adopt the AR-based try-on technology.

5.3 AR Immersion and Belief

Scholz and Smith [24] argues that to run successful AR campaigns, marketers need to carefully identify when and/or what triggers virtual content to be displayed and how they integrate with real objects. The authors argue that the maximum user-brand AR experience is primarily dependent on the AR affordance (i.e., availability of interactions and richness of the virtual content), which naturally depends on device processing speed, sophistication, and course cost. In other words, creating highly immersive AR environments depends on the cost of the technologies that make that environment. Also, marketers should aim to increase user-to-user engagement through creative AR environments where users can interact with each other while interacting with the brand content. The authors identify a third type of engagement: user-bystander engagement when AR users interact with other people in the physical environment but are not augmented themselves [25].

De Ruyter et al. [26] propose a contextual framework on the relationships between AR applications and advertising and marketing metrics. The authors identify three main constructs: context mapping, content matching, and customer experience, as drivers for advertising effectiveness metrics. Context mapping is the AR technology that recognizes and analyzes the environment, location, and objects users' see-through eye gaze tracking. Content matching is the process where AR embeds objects such as virtual images, text, or videos into the physical environment that users see in real-time. Users' belief and acceptance of those embedded objects rely on those embedded objects' quality and details. When context mapping and content matching happen successfully, AR customer experience is the third and most crucial component in AR. Here, the level of execution is strongly related to customers' feeling of spatial presence, which is users' feeling of being in a real (i.e., non-virtual) environment and "belief" of the VR environment [27].

5.4 Technology Involvement

Flavián et al. [28] classify the impact of virtual, augmented, and mixed realities on customers' core experience under four types of experience based on technology involvement and relevance to users' core experience. Directly supported experiences happen when AR technologies directly act in consumers' real world (e.g., testing how a new piece of furniture looks in a real-time room view). Indirectly supported experiences happen when AR technologies do not integrate content in consumers' real world (e.g., popping up information about a historical building while using smart glasses). AR technologies add relevant content to complement consumers' core experience (e.g., suggesting a piece of furniture and displaying it in a real-time room view). Finally, AR technologies create content that is not directly relevant to users' core experience in diverted-empowered experiences. Instead, it aims to divert them into another line of thought or goal (e.g., adding a restaurant menu within an apparel shopper's physical store view).

5.5 Public and Private Use

Rauschnabel [19] contends that consumers are motivated by AR technologies when they experience four different types of gratification. These are life efficiency (i.e., getting things done more easily), enjoyment, sensual (e.g., enhancing reality and comfort of wearing), social, and self-expression. However, the author differentiates between consumers' intention to use AR technologies in public and in private.

6 A 2020s Taxonomy for AR Customer Experience

The research on AR in customer experience is limited. While AR is proved to improve favorable customer variables, there is a need to examine how AR impacts the overall customer experience or journey [10]. Indeed, customer journeys are now more complex and interrelated than ever, considering multi-device usage and changes in customers' attitude toward advertising [6].

Earlier literature approached the taxonomy of AR applications in customer experience from an observation perspective (i.e., observing how some AR strategies impact certain variables or create certain types of experiences). While the latter approach is useful in understanding the recent phenomena, there is a critical need for a proactive approach to propose and guide future practice and research. The author also argues that current taxonomies are not in contact with the established theories of online customer experience and AR experiences in general. More importantly, earlier literature is insufficient to consider general industry trends such as the rise of social, video, and mobile use by customers.

Technology will continue to advance rapidly in this decade to alter the AR experience to become more believable, easy-to-use, and realistic. The author predicts that heavy headsets, goggles, and hazy 3D graphics will be replaced by mainstream microdevices worn or implanted into users' bodies. In other words, technology will no longer be the foundation for AR taxonomies as it will become mainstream, permanent, disruptive, cheap, consolidated, and simple. The disruptive smartphone revolution is a recent example.

The author proposes a novel approach to classify and integrate AR applications in marketing strategies. Guided by customer experience as a foundation, the new taxonomy attempts to link technology, industry trends, and literature on online customer experience in the last ten years [11, 20, 21].

Illustrated in Fig. 3, the author proposes a quadrant-based taxonomy with the X-axis of customer's "Spatial Presence" and customer's "Situation" (Y-axis). Spatial presence is the extent to which customers are emerged in the AR experience and feel it as real. In the next ten years, the author argues that AR experience's emergence will be more intense and closer to reality. Marketers and customers will be able to control the degree of that spatial presence or realism. The Y-axis of customers' "Situation" represents the type of situation that customers happen to be in. In micro situations, customers need to achieve an important short-term task, or

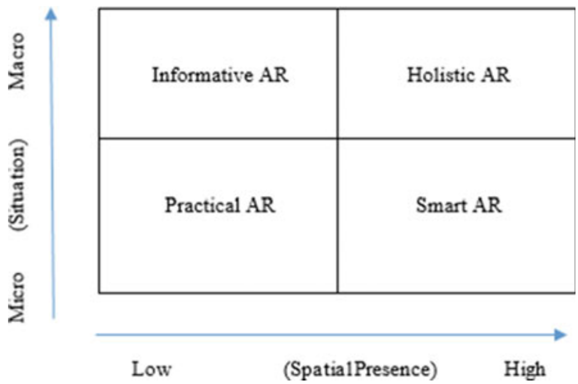


Fig. 3 A Proposed 2020s Taxonomy for AR Applications in Customer Experience

as Google named it, “micro-moments.” In Micro-moments, customers are more acquisitional and highly influenced by situational factors such as location, time constraints, and noise. Usually, these situations match low-involvement and convenient product shopping. In macro situations, customers have less time pressure and/or can execute longer-term and complex tasks. In macro situations, customers are more utilitarian, hedonic, and experiential types of shoppers. Usually, those situations match high-involvement, specialty products shopping.

The author proposes four main types of AR experience for marketers to integrate AR in their marketing strategies. Based on the proposed quadrant model illustrated in Graphical abstract, those types of experiences are Practical AR, Informative AR, Smart AR, and Holistic AR.

6.1 Practical AR Experience and Marketing Strategies

Customers are in a “Practical AR” experience exposed to or choose to be in a low spatial presence and are in a micro situation. Examples of those experiences are finding friends in a stadium or selecting a restaurant for lunch in a foreign country (Fig. 4). With such a limited timeframe and situational pressure to effectively and efficiently conduct those types of tasks, customers do not need to be fully immersed in an AR experience (i.e., high spatial presence). They need AR to help them get the right information quickly and accurately.

Practical AR experiences are highly dependent on wearable and implanted AR technologies such as smart glasses or implanted body chips. Customers still need to be an experienced reality and stay connected to it. AR here enhances the reality and improves the decision-making process. The brand experience here relies on the extent of how successful brands were in fulfilling customer’s urgent tasks smoothly and efficiently, eliminating clutter and noise, and predicting user’s requests. Artificial intelligence is critical in practical AR as it must analyze user’s shopping behavior, preference, and history to predict accurate shopping patterns.

Fig. 4 Practical AR experience. Adapted from mppglobal.com, 2020, https://www.mppglobal.com/news/blog/sports-optimising-the-fan-lifecycle/

When designing practical AR experiences, marketers should:

- Minimize customer experience time and maximize efficiency;
- Design highly customizable and personalized products;
- Integrate products and services in the AR virtual layer;
- Link content and promotion to wearable technology;
- Invest in linking the online (AR) to the offline experience;
- Have a limited product mix and product line widths;
- Follow an umbrella or family branding strategy;
- Adopt penetration, cost leadership, and competitive pricing strategies;
- Conduct sales promotions and limited time offers;
- Rely on local search engine optimization and location-based marketing;
- Adopt intensive distribution strategies and integrate physical points-of-sale in the AR experience;
- Rely on analytics to improve customers' experience; and
- Adopt rational/cognitive positioning and tone of voice.

6.2 Informative AR Experience and Marketing Strategies

The informative AR experience is when customers are in a low spatial presence and macro buying/shopping situations. Here, customers do not require to be highly immersed in an AR environment. They still need a great degree of reality where AR enhances and supports that reality. However, customers in that type of experience are involved in a long-term, routine, or high-involvement shopping situation where

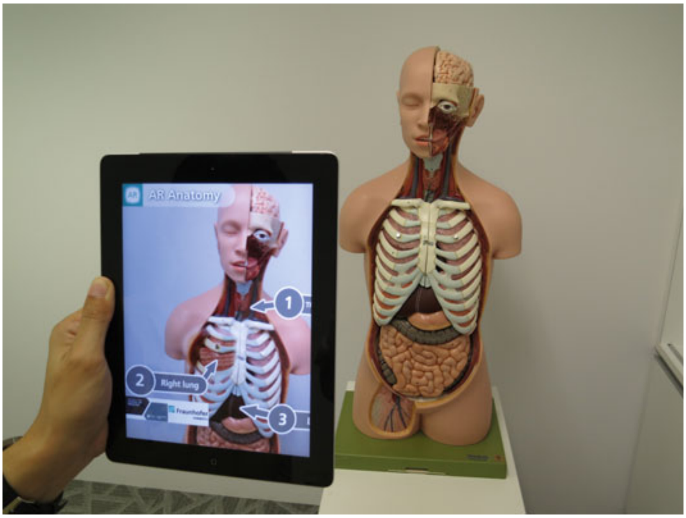


Fig. 5 Informative AR experience. Adapted from Wikipedia.org, 2020, https://commons.wikimedia.org/wiki/File:Augmented-reality-1957411_1920.jpg

they can afford to interact across multiple locations, devices, and times. An example of an informative AR experience is a student enrolled in an e-learning class (Fig. 5). He or she needs the most frequent and accessible content but does not necessarily require high degrees of immersion into virtual layers of content.

Multiple devices have provided an informative AR experience. The brand experience here depends highly on the consistency of experiences across different devices. Nowadays, experience consistency can be when users continue to watch a favorite Netflix movie on their TV from where they stopped watching on their mobile phone the night before. In AR, a consistent brand experience is much more complicated. A unified quality framework for AR content and affordance is essential whether customers see a hologram or are wearing a smart glass. AR technology should detect the appropriate tasks and content in relation to customers' current location and situation. AR here enhances reality continuously and consistently over time.

When designing informative AR experiences, marketers should:

- Unify customer experience across devices;
- Invest in content strategy and allow for content integration and growth;
- Drive habitual and frequent usage;
- Reward loyalty and referrals;
- Adopt subscription revenue models;
- Utilize promotion to drive product switching;
- Adopt executive distribution strategies;
- Have an extensive product mix and product line widths;

- Use brand and product line extension strategies;
- Adopt value-based pricing strategies; and
- Rely on analytics and user feedback to improve customers' experience.

6.3 Smart AR Experience and Marketing Strategies

Smart AR happens when customers have a micro-buying situation and a highly immersive (high spatial presence) AR experience. Here, customers are in a more virtual than a real environment where most of what they see, hear, touch, smell, and taste is virtual. However, the shopping situations they happen to be involved in are short-term, high-involvement, and acquisitional. An example of a smart AR experience is a customer attending the Olympics' opening ceremony in another country (Fig. 6). Here customers require the highest degree of a spatial presence, yet their shopping situation is occasional and short-term.

Customers in smart AR situations require short yet rich experiences. Smart AR experiences provided by a single device can entirely create an environment that looks as real and interactive as possible. In a smart AR experience, Scholz and Smith [24] model on AR engagement comes into play where user-user, user-brand, and user-nonparticipating user experiences are essential to have. Customers' first brand experience must be outstanding as it plays the most important determinant in generating future customer purchases [29].

Fig. 6 Smart AR experience. Adapted from singdaptive.com, 2020, https://blog.singdaptive.com/the-7-best-audition-tips-youll-ever-hear/

When designing smart AR experiences, marketers should:

- Carefully design customers' first experience with the brand;
- Integrate social strategies and encourage user-user interactions;
- Invest in new product development and innovation;
- Affiliate marketing, cross-selling, and co-branding;
- Adopt price skimming and niche pricing strategies;
- Facilitate AR compatibility and integration;
- Rely on customers' feedback to improve customers' experience; and
- Adopt exclusive distribution strategies.

6.4 Holistic AR Experience and Marketing Strategies

Holistic AR is the most futuristic type of experience. Here customers are always in a highly immersive AR environment and high-involvement, long-term macro shopping situation. An example of a holistic AR experience is a customer shopping for a new car (Fig. 7). Here, AR is integrated extensively across the entire customer decision-making process (i.e., need recognition, information search, alternatives evaluation, purchase, and post-purchase evaluation) [23]. That customer is exposed to targeted AI-based ads that motivate him or her to consider buying a new car via a permanent AR implanted chip or glasses. The customer then presents an entirely AR-based collection of cars with compassions, options to ask experts and friends, and prices. The customer can then be in an entire driving simulation that provides the look, sound, power, and even smell of each car shortlisted. The customer can interact with the dealer via AR, complete the transaction, and get the car. After buying, that customer can be exposed to branded self-driving technology, AR-enabled window solid, workspace, etc. Customer service is fully automated through bots and facial recognition (Fig. 7).

Customers in holistic AR are not just within a buying process. Holistic AR makes brands integral parts of customers' daily lives. Holistic AR experience is not limited to one situation, brand, or journey. It requires both single and multi-device AR technologies that enable customers to move from permanent (e.g., wearables) to temporary (e.g., flat surface or hologram workspace) smoothly and consistently. Holistic AR does not mean the full power of brands over consumers. The latter will be empowered to control what they are exposed to. Holistic AR represents a challenge to marketers as it represents an opportunity. Marketers must find innovative ways to add real value to consumers and respond to the increasingly demanded quality of products and services.

When designing holistic AR experiences, marketers should:

- Move from the concept of "brand touchpoints" to a single "point" that is integrated into customers' everyday activities (i.e., holistic experience);
- Move from the concept of content to experiences;

Fig. 7 Holistic AR experience. Adapted from automobile-propre.com, 2020, https://www.automobile-propre.com/bmw-i-visualiser-bmw-i3-i8-en-realite-augmentee/

- Expand product mix and offerings to include complementing solutions for customers' daily needs;
- Adopt integration and disintermediation business strategies;
- Rely on analytics and customers' feedback to improve customers' experience;
- Adopt selective or intensive distribution strategies;
- Invest in products comparisons and evaluation; and
- Adopt emotional/affective positioning and tone of voice.

7 Theoretical Implications and Future Research

The author proposes a novel theoretical model to classify the types of AR experiences from a marketing standpoint. The model links technology-based theories on AR experiences to marketing-specific shopping situations. It liberates online customer experience theories from the rational versus emotional debate that has been exhausted for the past three decades. The model builds on the knowledge of customer experience/journey [30], brand touchpoints (i.e., pre-purchase, purchase, and post-purchase) [21], and the recent buying micro-moments [16].

The author defends the notion that AR-enabled customer experience in the 2020s will no longer be before, during, and after purchase. They will be either short specific situations or continuous and integrated into consumers' daily lives. The author also defends that the classical consumer decision-making process [23] is still applicable in AR-enabled marketing strategies yet is more interconnected and complex.

The author encourages future researchers to test the proposed model from an empirical perspective and test its validity. More neuromarketing experiments are necessary to scientifically test customers' brain reactions when exposed to high spatial presence environments.

Augmented Reality (AR) will alter marketing forever. The author concludes this chapter with the following points:

- Proposed AR-enabled marketing taxonomy is four main types of customer experiences: Practical, Informative, Smart, and Holistic.
- Before 2030, customers may not differentiate what is real from what is not within an AR experience.
- Future marketing content will serve multiple customer realities (i.e., real, enhanced, favorable, and mixed).
- Marketing strategies should drive technology, not the other way.

8 Conclusion

Marketers should classify AR-enabled strategies based on customers' situations and experiences, not based on how AR technologies are classified. Also, they need to avoid approaching customer journey from a linear perspective such as Simon's [31] consumer decision-making process (i.e., need recognition, information search, evaluation, purchase, and post-purchase) or separated brand touchpoints before, during, and after purchase [21]. A marketing-based AR taxonomy should consider developing specific shopping experiences that utilize AR technologies and not the opposite.

The author proposes four main types of experiences that marketers should create for their customers based on spatial presence and customer shopping situation. These are Practical AR (low spatial presence and micro-moment), Informative AR (low spatial presence and macro-moment), Smart AR (high spatial presence and micro-moment), and Holistic AR (high spatial presence and macro-moment). By creating marketing strategies that understand those types of experiences, marketers can customize their marketing mixes to reach and retain customers across all types of experiences.

References

1. Abrar K (2018) Impact of augmented reality on consumer purchase intention with the mediating role of customer brand engagement : moderating role of interactivity in online shopping, Bahria University. J Manage Technol 1(2):64–80
2. Alimamy S, Deans K, Gnoth J (2018) An empirical investigation of augmented reality to reduce customer perceived risk. In: Marketing transformation: marketing practice in an ever changing world. Springer, Cham, pp 127–135. https://doi.org/10.1007/978-3-319-68750-6_42
3. Badkoube B et al (2019) Augmented reality for marketing, interactive advertising bureau. Available at: https://books.google.co.uk/books/about/Augmented_Reality_for_Developers.html?id=8xhKDwAAQBAJ&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false%0Ahttps://mediatum.ub.tum.de/1286695

4. Bayern M (2019) Internet of the senses is on the horizon, thanks to AR and VR—TechRepublic, TechRepublic. Available at: https://www.techrepublic.com/article/internet-of-the-senses-is-on-the-horizon-thanks-to-ar-and-vr/. Accessed 11 Aug 2020
5. Belsky S (2018) 16 fascinating augmented reality quotes everyone should read, Forbes. Available at: https://www.forbes.com/sites/bernardmarr/2018/09/05/16-fascinating-augmented-reality-quotes-everyone-should-read/#7522de015107. Accessed 9 July 2020
6. Bonetti F, Warnabyb G, Quinna L (2018) Augmented reality and virtual reality in physical and online retailing : a review. Synthesis Res Agenda
7. Chylinski M et al. (2020) Augmented reality marketing: A technology-enabled approach to situated customer experience, Australasian Marketing Journal (AMJ). Elsevier Australia. doi: https://doi.org/10.1016/j.ausmj.2020.04.004.
8. Edwards-Stewart A, Hoyt T, Reger GM (2016) Classifying different types of augmented reality technology Annu Rev Cyberther Telemed 14:199–202
9. Eigenraam AW et al (2018) A consumer-based taxonomy of digital customer engagement practices. J Interact Market 44:102–121. https://doi.org/10.1016/j.intmar.2018.07.002
10. Feng Y, Mueller B (2019) The State of augmented reality advertising around the globe: a multi-cultural content analysis. In: 2017 American academy of advertising global conference, pp 86–94
11. Hugues O, Fuchs P, Nannipieri O (2011) New augmented reality taxonomy: technologies and features of augmented environment. In: Handbook of augmented reality. Springer, New York, pp 47–63. https://doi.org/10.1007/978-1-4614-0064-6_2
12. Ramírez-Fernández C, Morán AL, Meza-Kubo V (2018) Towards a taxonomy of feedback factors affecting the user experience of augmented reality exposure therapy systems for small-animal phobias. UCAmI, 1252.https://doi.org/10.3390/proceedings2191252
13. Rose S et al (2012) Online customer experience in e-retailing: an empirical model of antecedents and outcomes. J Retail 88(2):308–322. https://doi.org/10.1016/j.jretai.2012.03.001
14. AbouElgheit E (2018) Understanding Egypt's emerging social shoppers. Middle East J Manage 5(3):207. https://doi.org/10.1504/mejm.2018.093611
15. Pham TSH, Ahammad MF (2017) Antecedents and consequences of online customer satisfaction: a holistic process perspective. Technol Forecast Soc Chang. https://doi.org/10.1016/j.techfore.2017.04.003
16. Makarov A (2020) 9 Augmented reality trends to watch in 2020: the future is here, MobiDev. Available at: https://mobidev.biz/blog/augmented-reality-future-trends-2018-2020. Accessed 11 Aug 2020
17. Naimark M (2018) VR/AR fundamentals—(3) Other Senses (Touch, Smell, Taste, Mind), Medium. Available at: https://medium.com/@michaelnaimark/vr-ar-fundamentals-3-other-senses-haptic-smell-taste-mind-e6d101d752da. Accessed 11 Aug 2020
18. Rauschnabel PA, Felix R, Hinsch C (2019) Augmented reality marketing: how mobile AR-apps can improve brands through inspiration. J Retail Consumer Serv 49:43–53. https://doi.org/10.1016/j.jretconser.2019.03.004
19. Rauschnabel PA (2018) Virtually enhancing the real world with augmented reality holograms : use and gratification perspective : an abstract virtually enhancing the real world with holograms : an exploration of expected gratifications of using augmented reality smart glasses.https://doi.org/10.1002/mar.21106
20. Yaoyuneyong G et al (2016) Augmented reality marketing: consumer preferences and attitudes toward hypermedia print ads. J Interact Advertis 16(1):16–30.https://doi.org/10.1080/15252019.2015.1125316
21. Javornik A (2016) Augmented reality: research agenda for studying the impact of its media characteristics on consumer behavior. J Retail Consumer Serv 30:252–261. https://doi.org/10.1016/j.jretconser.2016.02.004
22. Verdict (2020) History of virtual reality: timeline, Verdict.co.uk. Available at: https://www.verdict.co.uk/history-virtual-reality-timeline/. Accessed 9 July 2020

23. Yim MYC, Park SY (2019) I am not satisfied with my body, so I like augmented reality (AR) Consumer responses to AR-based product presentations. J Bus Res 100:581–589. https://doi.org/10.1016/j.jbusres.2018.10.041
24. Scholz J, Smith AN (2016) Augmented reality: designing immersive experiences that maximize consumer engagement. Bus Horizons 59(2):149–161. https://doi.org/10.1016/j.bushor.2015.10.003
25. Tsai S pei (2020) Augmented reality enhancing place satisfaction for heritage tourism marketing. In: Current issues in tourism. Routledge, pp 1078–1083. https://doi.org/10.1080/13683500.2019.1598950
26. de Ruyter K et al (2020) Seeing with the customer's eye: exploring the challenges and opportunities of AR advertising. J Advertis 49(2):109–124. https://doi.org/10.1080/00913367.2020.1740123
27. The Washington Post (2020) Fitting rooms are going virtual, The Washington Post. Available at: https://www.washingtonpost.com/business/2020/07/09/virtual-try-ons-are-replacing-fitting-rooms-during-pandemic/. Accessed 9 Aug 2020
28. Flavián C, Ibáñez-Sánchez S, Orús C (2019) The impact of virtual, augmented and mixed reality technologies on the customer experience J Bus Res 100:547–560.https://doi.org/10.1016/j.jbusres.2018.10.050
29. Google (2016) The basics of understanding key consumer micro-moments—think with Google. Available at: https://www.thinkwithgoogle.com/marketing-resources/micro-moments/micro-moments-understand-new-consumer-behavior/. Accessed 11 Aug 2020
30. Goebert C, Greenhalgh GP (2020) A new reality: fan perceptions of augmented reality readiness in sport marketing. Comput Hum Behav 106:106231.https://doi.org/10.1016/j.chb.2019.106231
31. Simon HA (1959) Theories of decision-making in economics and behavioral science. Am Econ Rev 49(3):253–283. https://doi.org/10.2307/1809901

Emad AbouElgheit Ph.D., is a passionate marketing and communications exponent with over 20 years of global experience in diversified sectors that include B2B software development, e-commerce, clean energy, healthcare, financial services, management consulting, and non-profit. Before academia, he held plenty of senior executive positions at international organizations across the Middle East. These include the United Nations, Gallup, and ITWorx. AbouElgheit has been leading various digital marketing, branding, creative, and communications strategies and initiatives throughout his career. Dr. AbouElgheit obtained his Ph.D. in International Business Management at the International School of Management (ISM) in Paris, France, in 2016. Currently, Dr. AbouElgheit is an Associate Professor of Marketing at Western Connecticut State University. He has several publications in marketing and digital consumer behavior. Also, Emad has been a presenter and moderator at various academic and professional conferences.

14 Can Viral Marketing Content Spread According to the Rumor Formula?

Silvia Sivera-Bello

For the world to receive information from the world and enjoy it, now computers and butterflies suffice.

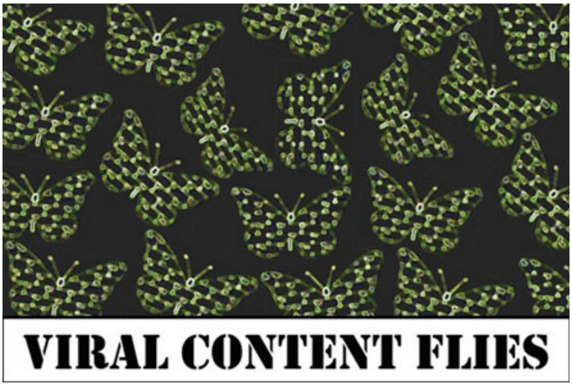
Italo Calvino [1]

Summary

On social media, brands use the potential of message virality to cut their investment budgets in traditional media and increase their audience. Users share marketing messages spontaneously, helping campaigns to spread exponentially. However, what type of content goes viral? In the so-called post-truth era, where fake news spreads faster than real news, it is worth analyzing whether something similar happens in the field of viral marketing. In an analysis of the content from 651 viral advertising videos, ambiguity, one of the essential factors in the Allport and Postman formula for spreading rumors, was detected in 42% of cases. According to these authors, a rumor spreads if the content is important and necessarily ambiguous. Based on their work, this study analyses key factors in a hypothetical viral advertising formula. These factors are ambiguous but also strategic creative drivers and advertising genres (viral strains), which together form a construct that replaces the factor importance in the original formula. The results could help guide content creators and brand managers in their communication strategies.

S. Sivera-Bello (✉)
Universitat Oberta de Catalunya, Barcelona, Spain
e-mail: ssivera@uoc.edu

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_14



Viral content flies.

The code of this chapter is 01010110 01101001 01110010 01100001 01101100 00100000 01101101 01100001 01110010 01101011 01100101 01110100 01101001 01101110 01100111.

1 Introduction

Two decades ago, viral marketing was injected metaphorically into the professional advertising world as a definitive vaccine for perceived audience immunity to marketing messages. Getting consumers to spread adverts organically among their contacts was seen as a panacea for an advertising market that had become over-saturated with information. However, this was not a new solution; word of mouth was already spreading recommendations between peers through so-called whisper campaigns documented in the 1930s [2].

Information and communication technologies facilitated an exponential leap toward viral advertising. Online devices enabled marketing conversations to flow among users over social media, either as recommendations or by merely forwarding advertisers' messages. It is this technology-enhanced macroscopic scaling up that fundamentally differentiates word of mouth advertising from viral advertising. Advertisers' messages can now reach larger audiences, more quickly, with better control and measurement, at low or virtually zero media cost and, above all, with greater reliability (as the recommended content is no longer as volatile as when spread by word of mouth and can be spread without modification).

Once digitalized, word of mouth (or word of mouse) facilitates virality at a *pandemic* dimension. However, two decades after what is considered the first viral marketing campaign, i.e., the launch of the Hotmail webmail service, professionals and academics are still asking questions about how viral advertising works. Why do some campaigns succeed online while others fail to reach a reproduction rate to ensure contagion? Where does the viral appeal of the marketing message lie? What do the most successful messages have in common? In short, is there a formula for virality?

The interest in addressing these issues is reflected in the professional and academic literature. In the first decade of research, the prevailing approach maintained that having an impact on the right opinion leaders (influencers) was more important than the viral advert content [3, 4].

However, more recent research raises doubts about this approach and suggests focusing on creating viral content is more effective than activating influencers [5–7], although the search continues for tools to help identify the social media nodes with the most influence [8].

2 From Word of Mouth to Rumor

The academic world-first showed an interest in viral marketing at the end of the twentieth century, based on the tradition of scientific study into word of mouth, which came out of the positivist school of communication research in the United States after the Second World War [2]. This approach recognizes the power of word of mouth to influence consumers and identifies three essential characteristics: i, interpersonal communication, not mass communication; ii, the communication of marketing content; and iii, marketing content not perceived as such.

There are references to word of mouth marketing tactics in the 1930s, led by *rumor peddlers*, but the first formal study in this field was headed by Whyte [9], who researched how the advantages of air conditioning units were spread by word of mouth in a neighborhood in Philadelphia.

Academic literature has studied both differences and similarities between word of mouth advertising and rumors. Thus, Rosen [10] differentiates word of mouth and rumor marketing (including buzz marketing) as viral marketing techniques. In other words, this author identifies differences between the two concepts, while Boquera [11] states that “rumor was no longer considered a harmful element for organizations and virtually became an advertising medium.” This author classifies rumor by use: planned (to promote a product, image, or public figure) and instrumental (to entertain, start a conversation, attract attention or contribute to group cohesion).

Paradoxically, it is in the field of communication where rumor has been least studied. Since Stern [12] first researched message distortion, rumor has mostly been studied in psychology, from a variety of perspectives: its influence [13], its functions [14, 15], its nature [16], and its categorization [17]. More recently, due to the proliferation of fake news in the digital media, similar studies have started about diffusion channels [18–21].

3 The Rumor Formula

During the Second World War, the US Government commissioned Gordon W. Allport, a psychology professor at Harvard University, to analyze the influence of rumor on civilian morale. With Robert H. Knapp, he founded the so-called *Rumor Clinic*, whose aim was to produce a weekly report in a local newspaper (*Boston Sunday Herald-Traveler*) presenting rumors circulating about the war and providing convincing rebuttals.

They established the key factors for spreading rumors in the formula:

$$\text{Rumor} = \text{Importance} \times \text{Ambiguity}(R = i \times a).$$

Thus, if the information has no importance or there is no ambiguity in the facts, there will be no rumor because if one of the factors equals zero, so will the final result [13].

Unlike psychology, which stresses each participant's individual needs, sociology sees rumor as a collectively formed element.

According to Shibutani [22], in this collective transaction, rumors are simultaneously formed and altered. He provides a holistic explanation for the rumor formation process, integrating into his explanation both how content is deformed and how it functions. He is critical of the Allport and Postman formula, as he believes the information does not necessarily has to be important, but it must be ambiguous.

Knopf [23] replaces importance with a problematic situation, while Chorus [24] adds a new variable: individual critical sense. The greater the critical sense (c), the fewer rumors will spread. Hence, $R = i \times a/c$.

Thus, the only concept questioned in the initial Allport and Postman formula is *importance*, while ambiguity is generally accepted, although more recent authors prefer the term *lack of evidence* [25].

There are diverging opinions on the factor *importance*, as different schools of thought prioritize different aspects, depending on their interests. Furthermore, it is difficult to define the importance of a message with scientific precision: where is the dividing line between an important topic and a topic of possible importance to a given public? In some cases, this might be clear, but in creative advertising content, it is considered a key factor (or the key factor) in triggering a viral process.

Morin [26] was one of the first sociologists to look beyond message deformation, concluding that rumors are full of cultural symbols. These contributions are particularly relevant as they make a direct connection between rumor and virus. Morin returns to the concept of rumors as disease, first mentioned by Allport and Postman in the *Rumor Clinics*, referring to them using terms such as *virus* or *germ* and describing how an antidote or death can stop them. He uses the term *drivers* to describe rumor transmitters and *antibodies* to refer to rumor paralyzers, while the phases through which rumors pass are given names such as *incubation*, *spread*, *metastasis,* and *traces.*

In the field of anthropology, research has looked into the extent to which rumors can cause racial conflict [27], while folklore studies examine rumors as urban legends [28, 29]. Using content analysis, Koenig [30] pioneered the study of rumors dealing exclusively with commercial products, looking beyond the three “Cs” (catastrophe, crisis, and conflict) to propose conspiracy and contamination myths as possible variables for the classification of rumors related to commercial products. Later on, Kapferer [31] looked into reasons why rumors spread and their particularities in different situations: business, politics, finance, the stock market, and gossip press. More recently, DiFonzo and Bordia [32] continue to approach rumors from social and organizational perspectives.

From a structural analysis approach, Gritti [33] provides an important contribution to the narrative of rumor, in complete contrast to Shibutani [22] regarding deformability: independently of how elaborate the rumor is, it always preserves a constant structure that is resistant to change. According to this fixed structure, rumors have a narrative development that features given characters within a specific time frame (mythical time)and different narrator positions.

4 A Formula for Virality

As there is little consensus regarding how the initial Allport and Postman formula might be revised, and given that its simplicity allows for its practical adaptation to the field of viral advertising, the following formula is proposed:

$$V = \mathrm{cc}(d + s) \times a$$

where advertising content virality (V) equals creative content (cc) produced by drivers (d) and viral strains (s), in which there needs to be a degree of ambiguity (a).

The factor *importance* in the initial formula for the spread of rumors is replaced by the sum of two variables that characterize the type of creative content used in viral video advertising.

The advertising literature has also failed to reach a consensus on key creative aspects in advertising (especially audiovisual adverts). Hence the two complementary elements discussed below are used to provide greater depth to the analysis.

4.1 Creative Drivers

Authors and researchers in advertising and virality have shown an interest in understanding the most strategic creative triggers, generally termed *drivers*.

Phelps et al. [34], in a study on word of mouth focusing on e-mail marketing, suggested that messages which produce strong emotions (humor, fear, sadness, or inspiration) are most likely to be shared. More recently, Berger [5] corroborated and developed upon this approach, adding another emotion to consider: awe, as a step beyond surprise.

Also, Dobele et al. [35] stated that successful viral marketing campaigns (in general terms, without specifying those operating via video) contain a seductive message due to the drivers of imagination, humor, or intrigue. Twose and Smith [36] agreed in stressing humor as a trigger for online marketing content virality and added the driver of involvement as developed in subsequent studies, such as Southgate et al. [37], who looked specifically at viral videos.

Gîrboveanu and Puiu [38] were categorical in stating that viral marketing's effectiveness is based on content quality. If it is interesting, entertaining, and informative (even better, if all three qualities are present simultaneously), it is sure to be shared. Furthermore, the authors specifically discussed the issue of marketing perception in virality: "*While it is viral, you do not want people to think its a virus*" [38].

Beyond this idea of perception, Cashmore [39] considered joy as the fundamental driver for virality: people share content because it makes them happy. Moreover, at the more pragmatic end, Berger and Milkman [40] stressed utility.

In line with most of the other authors cited, Southgate et al. [37] also agreed on the importance of humor but included the additional—and positive—determining factor: branding. Their work shows that the creative drivers of humor, involvement, and branding are positive predictors for viral viewing. They recognized that humor and involvement are more important and recommended them as the main mechanisms to consider with online videos; however, they also stressed that the positive correlation with branding is fascinating. Thus, they suggested a strong brand presence may help increase views, as it facilitates recall and searches for the video, should they be required.

About humor, they proposed the acronym laugh-out-loud funny, edgy, gripping, sexy (LEGS) to define the nature of entertainment and concluded that this type of content correlates positively to viewing figures. Finally, about involvement, they mentioned celebrity participation as another factor that tends to increase views. Notably, the authors placed greater value on the role of celebrities in the online rather than the offline context as a source of entertainment, gossip, and news.

Specific sector-based research [41–44] also considers factors such as useful information, novelty, content valence (whether it is positive or negative), emotion, and exemplification (the latter should be classified as a viral strain, given that it is approached as a narrative resource in the context of story-telling).

Thus, to summarize the above contributions, creative drivers fall into seven basic categories:

- Emotion: it aims to elicit feelings, both positive and negative. These may be primary or basic emotions (sadness, happiness, surprise, disgust, fear, rage), secondary or background emotions (discouragement, enthusiasm), and social emotions (embarrassment, gratitude, admiration, pride, jealousy, kindness, obfuscation, irritation);

- Sensation: it refers to messages based on sensory stimuli, whether visual (chromatic or achromatic), auditory, olfactory/gustatory, tactile, kinetic, or kinesthetic;
- Involvement: it attracts attention or awakes the public's altruism and commitment;
- Amusement: it adverts based on humor, jokes, and even comic violence [45];
- Information: the content limited to providing information on the product, service, or brand, serves as a reminder of its existence or possible uses;
- Utility: the content needs to be useful for users and adds value for products/services as an excuse or trigger for communication; and
- Rumor: it evokes ambiguous messages, open or vague narratives, or striking content to start up a viral communication or a public conversation.

4.2 Viral Strains

Numerous authors have identified and classified creative resources, genres or sub-genres in advertising while failing to produce a consensus taxonomy [46–55].

To clarify the minimum common denominators in the proposals made by these ten authors, the following list of creative resources is proposed and transposed to the field of virality, termed *viral strains*:

- Advertising strain: implying the presence—implicit or explicit—of a seller, including influencers, endorsements, presenters, satisfied users, and celebrities. This is most easily recognized as an advert due to its formal and persuasive features;
- Journalistic strain: this refers to formally the opposite of the above, as it avoids the appearance and language of conventional advertising. It uses supposedly objective and informative genres and formats from journalism, such as the report;
- Cinematic strain: it includes content inspired by any cinematic sub-genres and so-called slices of life and aspirational recreations;
- Televisual strain: these borrow formats and styles from television, especially those associated with entertainment (reality shows, quiz shows, or hidden cameras);
- Science-like strain: under the guise of scientific discourse, these recreate sociological or technological pseudo-experiments;
- Cartoon strain: these use comics and comic strips, cartoons, and digital creations;
- Fantasy strain: science fiction, magic, special effects, and even surrealism are used to create a dreamlike or fantasy discourse;
- Performance strain: this refers to dramatized content that might occur in the street (guerrilla actions), spontaneously or otherwise, generated by brands to impress the public;

- Rhetorical or conceptual strain: this covers items built using a rhetorical figure or conceptual resource from advertising;
- Comparative strain: this aims to establish the superiority of the advertised product or service over the competition or compare two specific associated situations;
- Demonstrative strain: this presents the product or service in use, usually with the classic problem–solution design;
- Musical strain: a tune or song forms the basis for communicating the message;
- Intrigue strain: this includes teaser adverts, with or without their subsequent resolution. It is initially the closest strain to the rumor driver;
- Interactive strain: this refers to content that requires user action to give the message meaning or demands explicit participation in other channels or platforms;
- Erotic strain: this uses sensuality and eroticism as its attraction;
- Esthetic strain: the clearest exponent of the artistic genre, in both static and moving images. The message is recreated in artistic form, even if this is *kitsch*; and
- Imitative strain: this includes me-too advertising, messages that try to emulate the category leader, and the creative path of swapping roles among the advert's characters.

To summarize, while viral strains refer to the type of creative content used, drivers refer to the motivation or strategic trigger used by advertisers to produce this content. Both elements of the formula are a construct related to the next factor.

4.3 Ambiguity

In 1999, the first advert whose viral strategy was based on ambiguity was launched. This was the teaser trailer for the psychological horror film *The Blair Witch Project* (written and directed by Daniel Myrick and Eduardo Sánchez), which obtained 75 million views in the campaign's first week. The campfire advertising agency designed a pre-release promotion, which suggested the story was real [56]. The story in question described the mysterious disappearance of three film students while recording a documentary in the woods in the US state of Maryland. The audiovisual material had supposedly happened a year later, providing a basis for event reconstruction. It was not a teaser designed as a conventional trailer. For this, the work's ambiguity hinged on doubts regarding the content itself and uncertainty for its diffusion.

Some years later, in 2004, the agency Crispin Porter + Bogusky designed an *advergame* (a game with advertising purposes) for Burger King as the center of a campaign to publicize the fast-food chain's new chicken burger. The game was hosted on the website www.subservientchicken.com and received over 15 million visits in the first week. Its viral appeal lay in the illusion of real-time interactivity perceived by users, who could enter a command (dance, jump, sleep, etc.) and make a character in a chicken costume perform the actions in front of a webcam in their lounge. In its design, the game covered a wide range of orders, thus creating reasonable doubt about the simultaneity of the character and users' commands. Ambiguity was also produced by the cryptic presence of the brand on the website (a discreet link under the title "BK Tendercrisp"). This case highlights the early influence of word of mouth on viral marketing, as the aim was to avoid videos being perceived as adverts.

In these and the previous examples, three levels of ambiguity can be detected in viral advertising videos: the ambiguity of authorship, the ambiguity of content, and double ambiguity.

4.3.1 Ambiguity of Authorship

Doubts swirl about one who is responsible for the work, either because an advertiser does not sign the video or because user-generated content could replace the marketing intent. Occasionally the brand might appear camouflaged in certain shots, thus generating conversations among users based on speculation and attempts to clear up doubts. This is the case with Ray-Ban's "Catch Glasses" viral advertising video, in which two people play with their sunglasses, but with no mention of the brand or close-up of the product. Only in one sequence could the brand slogan ("Never hide") be read, discreetly, and unconventionally integrated into the scene (Fig. 1).

4.3.2 Ambiguity of Content

Uncertainty is common regarding the truthfulness of the narrative. This usually involves blurring the limits between the advertising fiction and reality (about the characters and/or the action involved). An example illustrating this characteristic is the video "The Epic Split Feat" by Volvo Trucks, in which the actor Jean Claude Van Damme carries out a feat of acrobatics between two moving lorries (Fig. 2).

Fig. 1 Still from the Ray-Ban viral advertising video "Catch Glasses." The brand slogan appears written on the car window

Fig. 2 Still from the Volvo Trucks viral advertising video "The Epic Split Feat"

4.3.3 Double Ambiguity (Authorship and Content)

It is the case of the viral video "Guys backflip into jeans" by Levi's clothing brand, where the brand name never appears in close-up (a simple deduction that not all the jeans appearing in action are necessarily Levi's; hence it is not an advertising message) and that the acrobatics performed to get into the jeans are plausible though not necessarily real (Fig. 3).

Fig. 3 Still of Levi's viral advertising video "Guys backflip into jeans"

5 Testing the Formula

It is the time to test with data from a set of highly representative, homogeneous, and relevant viral advertising videos that are as neutral as possible (their selection is not based on criteria established by a jury, professional, or researcher). The sample was compiled over 42 months by activating a Google alert system with the keywords "viral marketing" (in English and Spanish), which provided a total of 12,610 hits on the topic.

From this content, references to videos posted on YouTube (strategically designed to go viral) were extracted, and the definitive study corpus was reduced to 651 viral advertising videos, thereby ensuring the homogeneity and relevance of the sample. Bearing in mind that viral advertising videos can be considered an infinite population, the sampling error was ±3.8% ($p = q = 0.05$ at a 95% confidence interval) to ensure representativity.

Data analysis using contingency tables permitted the different bivariate qualities of the sample to be described quantitatively (as frequencies and percentages) and the variable pairs' independence to be tested, using the χ^2 statistic.

The independence test proposes, as a null hypothesis, that the classification criteria are independent, i.e., a video belonging to one class of classification variable (such as ambiguity) does not affect its likelihood of belonging to any of the other variable types (such as drivers or strains).

The quantitative analysis was based on the frequencies at which certain indicators appear (*prevalence*, in the viral metaphorical language) but also considered the potential significance of isolated presence and even absence of certain elements. The results for each element in the formula are given below.

5.1 Drivers

The predominant creative driver in the sample was humor, matching results obtained by authors in previous research [35–38, 57].

The second prevalent driver is a rumor, which arises from ambiguous messages and open, vague, or astonishing content. It should be remembered that only two teams of authors have discussed this concept in the literature reviewed. Dobele et al. [35, 57] considered intrigue as one of the main triggers for creative viral content (together with emotion, entertainment, and imagination), while Southgate et al. [37] referred to certain forms of gossip news to justify brands' use of celebrities.

The third driver detected, in order of prevalence, is a sensation, which aims to create interest through sensory stimuli, and the fourth is emotion. Negative emotions were detected in 16 out of a total of 92 videos identified as using this driver, thereby calling into question suggestions that only positivity will make advertising go viral [58].

Brand or service was the fifth-placed driver for launching a viral video. Gîrboveanu and Puiu [38] are the standard authors about this driver, which has

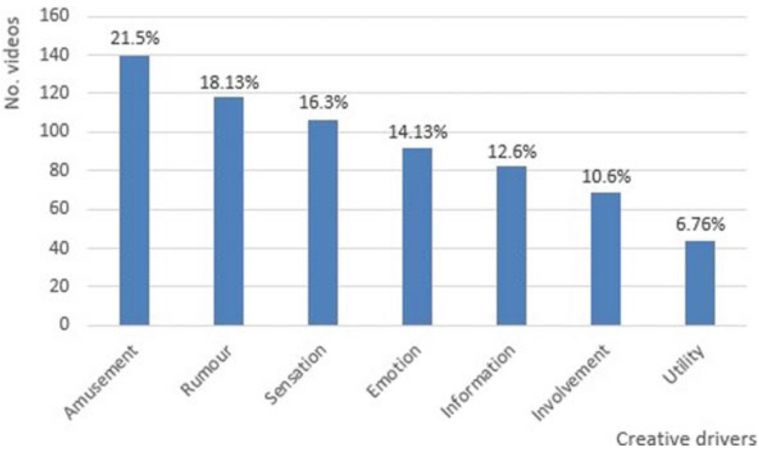


Fig. 4 Prevalence of creative drivers in the videos analyzed

historically been one of the primary communication goals (together with branding and call-to-action).

The penultimate driver is involved. Twose and Smith [36] and Southgate et al. [37] detected this creative trigger in their research, recently and eloquently expressed in the phrase "When we care, we share" [6]. As is to be expected, 68% of videos in this category are by advertisers in the service sector and institutions, as most are promoted by NGOs or non-profit foundations seeking to involve society in their cause.

In the last place in terms of prevalence is the utility driver, despite Berger and Milkman [40] identifying it as a determining factor for sharing online content.

Figure 4 shows the simplified driver frequency extracted from the data analysis (percentages in descending order).

5.2 Viral Strains

The most eloquent conclusion from the analysis of viral strains is the confirmation that viral advertising videos are, above all, adverts. In other words, they are identifiable as marketing artifacts.

This statement is based on the fact that the most commonly used strains are most closely associated with persuasive marketing language: rhetoric or conceptual and classic advertising strains. In other words, viral advertising videos mainly adopt resources commonly used in televisions ads: they use rhetorical figures as the conceptual pillar for communication (10.75% of cases) or turn to different types of influencers and advertising resources that make videos recognizable as adverts (10.29% of the cases analyzed). In total, over 21% of the videos in the sample match the genre style categories most closely related to advertising.

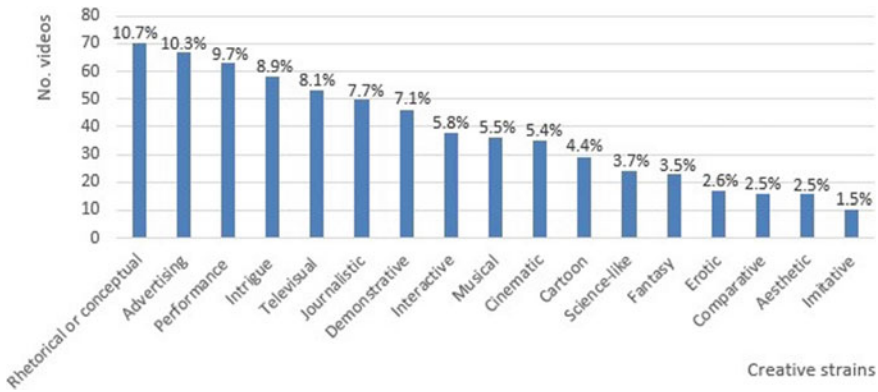


Fig. 5 Prevalence of creative strains in the videos analyzed

The third most prominent strain is performance. Some advertisers (such as Coca-Cola, Heineken, Levi's, Peugeot, Samsung, T-Mobile, and TNT) repeatedly use this, although advertisers in the technology sector are the most frequent users of this strain.

The fourth strain is intriguing, which includes teaser messages and tries to persuade viewers to take further action to solve the mystery. As might be expected, this strain is most frequently related to the rumor driver (58.62% of matches).

The fifth most prevalent viral strain is the televisual strain, which borrows formats and styles from television. Specifically, the use of the hidden camera format is particularly common (69.81% of the total for the televisual strain).

Figure 5 includes data for the strains.

5.3 Ambiguity

Ambiguity was analyzed using a dichotomous scale, considering content to be ambiguous if, as the definition of the adjective states, it is "*capable of two or more possible interpretations, and is usually intended to be so for the purpose of mystifying or confusing*" [59].

Among the 651 videos analyzed, a degree of ambiguity was detected in 276 (42.4%) cases. Therefore, it may be stated that ambiguity is not strictly a necessary condition for advertising virality. Consequently, the proposed hypothetical formula should be revised by changing the factors to addends:

$$V = \mathrm{cc}(d + s) + a$$

According to this new formulation, viral advertising content requires creative content (preferably based on specific drivers and viral strains) and a degree of ambiguity. Switching ambiguity from factor to addend shows that it is not an essential condition. Otherwise, a value of zero for ambiguity would mean there was no virality.

The statistical analysis of the viral advertising videos in the sample shows content to be the prevalent type of ambiguity. In total, 140 such cases were identified, representing 50.73% of the total. In second place are videos that use double ambiguity, both content and authorship. There are 71 cases in this category, representing 25.72% of the sample. Finally, close behind are ambiguous videos only for authorship, totaling 65, or 23.55% of the total. This latter figure contradicts specific authors' opinion, who consider there to be a greater likelihood of sharing viral videos that do not identify the advertiser and are therefore not associated with strengthening the brand [60]. Although there are numerous examples where a brand does not, apparently, feature in the commercial content, it is worth noting that some authors stress that branding is one of the primary goals of viral advertising [37, 61, 62].

Data analysis was used to produce contingency tables for the advertisers' sectors and content ambiguity. Significant correlations were found for only two sectors: 34.8% of videos in the consumer goods sector and 48.6% in technology, electronics, and telecommunications are ambiguous.

5.4 Relevant Correlations

Given that a single viral message may be based on more than one creative strain, the videos' analysis only identified the predominant one. To compensate for how this might limit the information, the data were crossed to detect the prominent creative drivers. By doing this, several significant correlations can be detected, which may be of interest to the academic and professional creative and viral advertising community. The basic matrix offers 119 possible driver and creative strain combinations (Table 1).

Having analyzed the significance of the correlations, the conclusion drawn from the data match is that the most relevant pairing in the set of variables analyzed is: emotion driver + rhetorical or conceptual strain.

The ambiguity variable is also analyzed about specific involvement in creative drivers and strains. Moreover, as expected, the strain most frequently related to the rumor driver is intriguing (58.62% of matches).

About drivers, ambiguity is significant in all of them, except humor. In other words, any trigger (except humor) implies that the content may contain a significant degree of ambiguity. Although this is the dominant driver (21.5% of the total), only 8.14% is ambiguous; hence it is not significant (0.129 according to the contingency table).

By contrast, the second most prevalent driver shows very high significance at the 99% confidence interval. In 94.9% of cases, rumor-based viral videos are ambiguous, compared to 30.8% of those considered ambiguous but not associated with this driver (Table 2). As this is a logical and predictable result, it is interesting to analyze cases that do not match this pattern. For instance, Gucci launched a video

Table 1 Summary of the cross between creative strains and drivers

Strains/creative drivers		Emotion	Sensation	Involvement	Amusement	Information	Utility	Rumour	Total
Advertising	Count	7	7	3	25	9	4	11	67
	% in drivers	7.7%	6.6%	4.3%	17.9%	11.%	9.1%	9.4%	10.1%
Journalistic	Count	6	3	7	6	6	12	10	50
	% in drivers	6.6%	2.8%	10.%	4.3%	7.3%	27.3%	8.5%	7.7%
Cinematic	Count	8	6	1	5	8	2	5	35
	% in drivers	8.8%	5.7%	1.4%	3.6%	9.8%	4.5%	4.3%	5.4%
Televisual	Count	12	4	8	18	4	2	5	53
	% in drivers	13.2%	3.8%	11.4%	12.9%	4.9%	4.5%	4.3%	8.1%
Science-like	Count	2	6	1	3	4	0	8	24
	% in drivers	2.2%	5.7%	1.4%	2.1%	4.9%	0.%	6.8%	3.7%
Cartoon	Count	0	8	3	4	11	2	1	29
	% in drivers	0.%	7.5%	4.3%	2.9%	13.4%	4.5%	0.9%	4.5%
Fantasy	Count	0	4	0	4	5	0	10	23
	% in drivers	0.%	3.8%	0.%	2.9%	6.1%	0.%	8.5%	3.5%
Performance	Count	17	14	3	12	3	1	13	63
	% in drivers	18.7%	13.2%	4.3%	8.6%	3.7%	2.3%	11.1%	9.7%
Rhetorical/conceptual	Count	18	6	14	21	9	1	2	70
	% in drivers	19.8%	5.7%	20.%	15.%	11.%	2.3%	1.7%	10.9%
Comparative	Count	5	1	2	1	5	0	2	16
	% in drivers	5.5%	0.9%	2.9%	0.7%	6.1%	0.%	1.7%	2.5%
Demonstrative	Count	8	6	4	9	9	7	3	46
	% in drivers	8.8%	5.7%	5.7%	6.4%	11.%	15.9%	2.6%	7.1%
Musical	Count	4	9	1	7	3	2	10	36
	% in drivers	4.4%	8.5%	1.4%	5.%	3.7%	4.5%	8.5%	5.5%

(continued)

Table 1 (continued)

Strains/creative drivers		Emotion	Sensation	Involvement	Amusement	Information	Utility	Rumour	Total
Intrigue	Count	2	6	3	8	2	3	33	58
	% in drivers	2.2%	5.7%	4.3%	5.7%	2.4%	6.8%	28.2%	8.9%
Interactive	Count	1	5	14	9	1	7	1	38
	% in drivers	1.1%	4.7%	20.%	6.4%	1.2%	15.9%	0.9%	5.8%
Erotic	Count	0	11	1	4	0	0	1	17
	% in drivers	0.%	10.4%	1.4%	2.9%	0.%	0.%	0.9%	2.6%
Aesthetic	Count	1	9	2	0	3	0	1	16
	% in drivers	1.1%	8.5%	2.9%	0.%	3.7%	0.%	0.9%	2.5%
Imitative	Count	0	1	3	4	0	1	1	10
	% in drivers	0.%	0.9%	4.3%	2.9%	0.%	2.3%	0.9%	1.5%
Total	Count	92	106	69	140	82	44	118	651
	% in drivers	100.%	100.%	100.%	100.%	100.%	100.%	100.%	100.%

Table 2 Contingency and chi-square test for ambiguity and the rumor driver

Contingency table			Ambiguity		Total
			No	Yes	
Rumour driver	No	Count	369	164	533
		% in the rumour driver	69.2%	30.8%	100.0%
	Yes	Count	6	112	118
		% in the rumour driver	5.1%	94.9%	100.0%
Total		Count	375	276	651
		% in the rumour driver	57.6%	42.4%	100.0%
Chi-square test		Value	Asymptotic sig. (2-sided)	Exact sig. (2-sided)	Exact sig. (1-sided)
Pearson's chi-square		162.776	.000		
Continuity correction		160.160	.000		
Likelihood ratio		181.947	.000		
Fisher's exact test				.000	.000
Linear-by-linear association		162.526	.000		
N of valid cases		651			

teaser directed by the filmmaker Frank Miller. The video aroused expectations around its premiere during the MTV Music Awards and, above all, created a surprising connection between the brand and Miller's esthetics to generate conversations. Thus, Gucci attempted to create a conversation about the brand through artistic provocation, but with added uncertainty.

The involvement driver also correlates to ambiguity, with significance at a 99% confidence interval. However, it is worth noting that this correlation applies to almost half the rumor-based videos. It is closely followed by the emotional driver, but in this case, significance means there is a lower tendency to ambiguity in 26.4% of cases than with any other driver (45%).

A similar correlation occurs with the sensation driver, which involves less ambiguity (34.9%) than in cases based on other drivers (43.9%). Finally, the information and utility drivers also show significance with ambiguity at the 95% confidence interval.

About viral strains, significant correlations with ambiguity are found in only 5 cases (out of 17 strains). From higher to lower significance, these are musical (p = 0.000), cartoon (p = 0.003), demonstrative (p = 0.006), esthetic (p = 0.042), and intrigue (p = 0.058). Atomization across such a broad set of categories might explain the low level summarized in the contingency tables.

Among the strains showing no significant correlation with ambiguity are the imitative, science-like, rhetorical, or conceptual and advertising strains. Consequently, no definite conclusions can be drawn from these results regarding ambiguity as a significant factor about the prevalent strains (rhetorical and advertising), although it is a significant factor for the dominant driver (emotion).

Finally, the analysis was fine-tuned by looking at the time factor analyzed to see if it impacted the correlations as mentioned above. To do this, data were crossed with two different groups: firstly, the videos longer than the sample average (2 min and 31 s); and secondly, videos shorter than average.

In general terms, the video's length does not alter the significance correlation compared to the correlation for the whole sample, and when it does change, this is generally for longer than average videos.

Referring specifically to drivers, longer videos appear to influence emotion, sensation, and involvement. However, in humor, information, utility, and rumor, the video's length does not alter the significance of ambiguity concerning the sample total.

In the case of viral strains, there are alterations in the following cases: cinematic, televisual, cartoon, performance, intrigue, and esthetic. Again, this is always for videos longer than 2 min and 31 s.

6 Conclusion

Analysis of 651 viral advertising videos shows that content based on emotions (positive or negative), explained using habitual advertising rhetoric (or conceptual) resources, shows greater significance and prevalence.

Ambiguity is present in almost half the sample analyzed and, due to the significant correlations with strains and drivers, it may be concluded that this is a condition worth considering, though not essential, for advertising virality.

Humor—in all its variants, including comic violence—is the predominant trigger but is not significant about ambiguous content. However, the rumor is close behind in second place; hence the role of ambiguity in viral content should not be underestimated.

Unlike previous studies [38, 63], where the study sample was considerably smaller than the one used here, this study finds that advertising virality is not indicated exclusively for advertisers in specific sectors. On the contrary, it is fairly widely distributed, with companies from all kinds of sectors placing strategic importance on viral advertising possibilities. Nevertheless, except for Google, the large offline advertising brands again rank highly in viral marketing.

Nor can it be stated that creativity is reserved to advertisers in specific business sectors. This is one of the collateral biases this statistical study of videos disproves, along with the idea that adverts need to be short to ensure they go viral (as the average length of the videos analyzed is 2 min and 31 s).

Besides the evidence from the discourse analysis, certain other relevant aspects have been detected that were not initially anticipated, such as the presence of viral strains not initially considered, e.g., digital sub-genres. In the performance strain, a considerable content group based on overcoming challenges was identified. Moreover, given the popularity of viral challenges in recent years, it may be worth tracking viral challenges associated with brands.

Interactivity, above and beyond forwarding or spreading advertising content, was detected as a viral strain in 5.84% of the sample. This could be interpreted in two ways: either there is a long way to go before the opportunities offered by ICT are fully used, or advertisers still doubt the Internet's potential as a participative medium.

Thus, we may say that ambiguity is a favorable condition for virality, especially when combined with the rumor and emotion drivers and the other triggers analyzed (except for humor). Ambiguity covers a wide terrain for an audience that considers certain marketing messages irresistible, although they occasionally fail to identify them as such. Metaphorically, referring to the quote at the start of this chapter, we could say that enjoying butterflies does not mean catching them.

Core Messages

- 651 viral videos analyzed.
- Ambiguity occurred in 42.4% of cases.
- Humor was the predominant trigger, but the most relevant pairing in the variables was emotion driver + rhetorical or conceptual strain (advertising genre).
- Ambiguity is an additional value in a hypothetical formula for advertising virality, although it is not essential to contagion.
- The findings led us to revise the initial formula by changing the factors to addends: **V**irality = **c**reative **c**ontent (**d**rivers + **s**trains) + **a**mbiguity.

References

1. Calvino I (2004) El castillo de los destinos cruzados. Siruela, Madrid
2. Kirby J, Marsden P (2006) Connected marketing: the viral, buzz and word of mouth revolution. Butterworth-Heinemann, Oxford
3. Adamic L, Adar E (2005) How to search a social network. Soc Netw 27:187–203
4. Kozinets R-V (2006) Click to connect: netnography and tribal advertising. J Advert Res 46 (3):279–288
5. Bakshy E, Hofman J-M, Mason W-A, Watts DJ (2011) Everyone's an influencer: quantifying influence on Twitter. In: Proceedings of the 4th international conference on web search and data mining, pp 65–74
6. Berger J (2013) Contagious: why things catch on. Simon & Schuster, New York

7. Watts D-J, Peretti J, Frumin M (2007) Viral marketing for the real world. Harv Bus Rev 85 (5):22–30
8. Sheikhahmadi A, Nematbakhsh M-A (2016) Identification of multi-spreader users in social networks for viral marketing. J Inf Sci 43(3):412–423
9. Whyte W-H Jr (1954) The web of word of mouth. Fortune 50:140–143
10. Rosen E (2000) The anatomy of buzz. Currency Doubleday, New York
11. Boquera E (2003) La informació ambígua: una introducció a les funcions i usos del rumor. Universitat Ramon Llull, Barcelona
12. Stern L-W (1902) Zur Psychologie der Aussage. Experimentelle Untersuchungen über Erinnerungstreue. Zeitschrift für die gesamte Strafechtswissenschaft 22(2/3)
13. Allport G-W, Postman L (1973) Psicología del rumor. Psique, Buenos Aires
14. Back K, Festinger L, Hymovitch B, Kelley H, Schachter S, Thibaut J (1950) The methodology of studying rumor transmission. Hum Relat 3(3):307–312
15. Festinger L (1950) Informal social communication. Psychol Rev 57(5):271–282
16. Kirkpatrick C (1932) A tentative study in expermimental social psychology. Am J Sociol 38 (2):194–206
17. Knapp R-H (1944) A psychology of rumor. Public Opin Q 8(1):22–37
18. Rochlin N (2017) Fake news: belief in post-truth. Libr hi tech 35(3):386–392
19. Rubin V-L (2017) Deception detection and rumor debunking for social media. In: The SAGE handbook of social media research methods, pp 342–364
20. Burkhardt J-M (2017) Combatting fake news in the digital age. Libr Technol Rep 53:8
21. Vosougui S, Roy D, Aral S (2018) The spread of true and false news online. Science 359 (6380):1146–1151
22. Shibutani T (1966) Improvised news: a sociological study of rumor. Bobbs Merrill, Indianapolis
23. Knopf T-A (1975) Rumors, race and riots. Transction Books, New Brunswick
24. Chorus A (1953) The basic law of rumor. J Abnorm Soc Psychol 48(2):313–314
25. DiFonzo N (2019) A Christian psychology of rumor. J Psychol Christ 38(1):3–21
26. Morin E (1973) La rumeur d'Órleans. Éditions du Seuil, París
27. Turner P-A (1993) I heard it through the grapevine. Rumor in African-American culture. University of California Press, Berkeley
28. Renard J-B (1999) Rumeurs et legends urbaines. Presses Universitaires de France, París
29. Dorson R-M (1999) Modern folklore. In: American folklore. University of Chicago Press, Chicago
30. Koenig F (1985) Rumor in the marketplace. The social psychology of commercial hearsay. Auburn House Publishing Company, Dover
31. Kapferer J-N (1989) Rumores. El medio de difusión más antiguo del mundo. Plaza&Janés, Barcelona
32. DiFonzo N, Bordia P (2007) Rumor psychology: social and organizational approaches. American Psychological Association, Washington DC
33. Gritti J (1978) Elle court, elle court, la rumeur. Stanké, Otawa
34. Phelps J-E, Lewis R, Mobilio L, Perry D, Raman N (2004) Viral marketing or electronic word-of-mouth advertising: examining consumer responses and motivations to pass along e-mail. J Advert Res 44(4):333–348
35. Dobele A, Lindgreen A, Beverland M, Vanhamme J, Van Wijk R (2007) Why pass on viral messages? Because they connect emotionally. Bus Horiz 50:291–300
36. Twose D, Smith D (2007) How effectively can ad research predict sales? Admap 487:42–44
37. Southgate D, Westoby N, Page G (2010) Creative determinants of viral video viewing. Int J Advert 29(3):349–368
38. Gîrboveanu S-R, Puiu S (2008) Viral Market. Ann Univ Petroşani Econ 8(1):223–230
39. Cashmore P (2009) YouTube: why do we watch? CNN.com http://edition.cnn.com/2009/TECH/12/17/cashmore.youtube/index.html
40. Berger J, Milkman K-L (2012) What makes online content viral? J Mark Res 49:192–205

41. Kim H-S (2015) Attracting views and going viral: how message features and news-sharing channels affect health news diffusion. J Commun 65:512–534
42. Cappella J-N, Kim H-S, Albarracín D (2015) Selection and transmission processes for information in the emerging media environment: psycholocial motives and message characteristics. Media Psychol 18(3):396–424
43. Kim H-S, Lee S, Cappella J-N, Vera L, Emery S (2013) Content characteristics driving the diffusion of antismoking messages: implications for cancer prevention in the emerging public communication environment. J Natl Cancer Inst Monogr 2013:182–187
44. O'Keefe D-J (2003) Message properties, mediating states, and manipulation checks: claims, evidence, and data analysis in experimental persuasive message effects research. Commun Theory 13:251–274
45. Brown M, Bhadury R, Pope N (2010) The impact of comedic violence on viral advertising effectiveness. J Advert 39(1):49
46. White H (1981) How to produce an effective TV commercial. Crain Books, Chicago
47. Ogilvy D (1983) Ogilvy & la publicidad. Folio, Barcelona
48. Baldwin H (1989) How to create effective TV commercials. NTC Business Books, Chicago
49. Ortega E (1992) La publicidad en televisión. Delphi-EAP-Mundiprensa, Madrid
50. Bassat L (2001) El libro rojo de la publicidad. Plaza & Janés, Barcelona
51. Rey J (1996) Palabras para vender, palabras para soñar: introducción a la redacción publicitaria. Paidós, Barcelona
52. León J-L (1996) Los efectos de la publicidad. Ariel Comunicación, Barcelona
53. Moreno I (2003) Narrativa audiovisual publicitaria. Paidós, Barcelona
54. Altarriba M (2005) Què dir, a qui i per què. Retòrica i redacció publicitària. Trípodos, Barcelona
55. Martínez-Sáez J (2009) Hacia una taxonomía del audiovisual publicitario. BOCC Biblioteca on-line de ciencias da comunicaçao, Universidade de Beira Interior, bocc.ubi.pt/pag/bocc-matinez-taxionomia.pdf
56. Del Pino C (2007) Nueva era en la comunicación comercial audiovisual: el marketing viral. Pensar la publicidad 1(2):63–76
57. Dobele A, Toleman D, Beverland M (2005) Controlled infection! Spreading the brand message through viral marketing. Bus Horiz 48(2):143–149
58. Eckler P, Bolls P (2011) Spreading the virus: emotional tone of viral advertising and its effect on forwarding intentions and attitudes. J Interact Advert 11(2):1–11
59. Rhvp (1996) Webster's encyclopaedic unabridged dictionary. Random House Value Publications
60. Sanagustín E (2009) Del 1.0 al 2.0: claves para entender el nuevo marketing. https://app.box.com/shared/tgoujqjm72
61. Golan G-J, Zaidner L (2008) Creative strategies in viral advertising: an application of Taylor's six-segment message strategy wheel. J Comput Mediat Commun 13(4):959–972
62. Porter L, Golan G-J (2006) From subservient chickens to brawny men: a comparison of viral advertising to television advertising. J Interact Advert 6(2):30–38
63. Woerndl M, Papagiannidis S, Bourlakis M, Li F (2008) Internet-induced marketing techniques: critical factors in viral marketing campaigns. Int J Bus Sci Appl Manage 3 (1):33–45
64. Boase J, Wellman B (2001) A plague of viruses: biological, computer and marketing. Curr Sociol 49(6):39–55

Silvia Sivera-Bello is Associate Dean for Emerging Programmes at the Faculty of Information and Communication Sciences at the Universitat Oberta de Catalunya (UOC). She also directs the UOC's Master's Degree in Strategy and Creativity in Advertising and teaches courses related to creativity and advertising within the University's bachelor's degree programs in communication and Digital Design and Creation. She is a researcher with the Learning, Media and Entertainment Research Group (GAME), recognized by the Catalan University and Research Grant Management Agency (AGAUR). In 2015 she was awarded the 6th Prat Gaballí Prize for Promoting Research into Advertising and Public Relations for her doctoral thesis on viral marketing. In 2020 the government of Catalonia awarded her the Jaume Vicens Vives Distinction for University Quality. She holds a degree in Information Sciences from the Universitat Autònoma de Barcelona (UAB), a master's degree in the Information and Knowledge Society from the UOC, and a Ph.D. from Ramon Llull University (URL).

Check for updates

Exaptation as a Design Strategy for Resilient Communities

15

Alessandro Melis and Telmo Pievani

> In King's College Chapel in Cambridge, for example, the spaces contain bosses alternately embellished with the Tudor rose and portcullis. In a sense, this design represents an "adaptation," but the architectural constraint is clearly primary. The spaces arise as a necessary by-product of fan vaulting; their appropriate use is a secondary effect. Anyone who tried to argue that the structure exists because the alternation of rose and portcullis makes so much sense [...]. Yet evolutionary biologists, in their tendency to focus exclusively on immediate adaptation to local conditions, do tend to ignore architectural constraints and perform just such an inversion of explanation. *[1]*

Summary

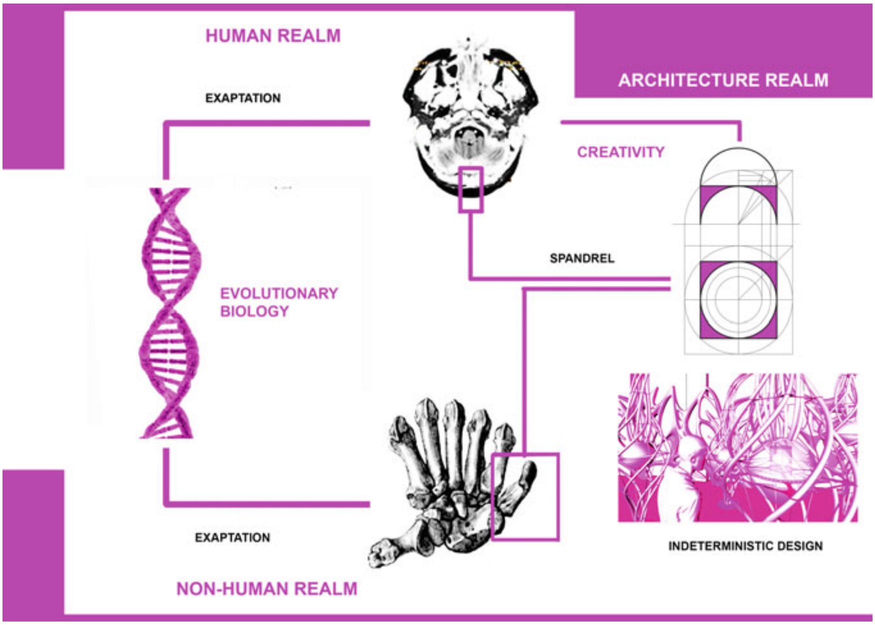
The chapter focuses on the significance and the originality of the study of the exaptation as a possibility to extend the architectural design toward more sustainable approaches aimed at enforcing urban resilience. The use of exaptation's definition in architecture corroborates the heuristic value of the cross-disciplinary studies on biology and architecture, which seems even more relevant in times of global environmental crises. Exaptation will be described, in the chapter, as a functional shift of a structure that already had a prior but different function. In architecture, a functional shift of a structure that already had a function may apply to forms of decorations embedded in architectural

A. Melis (✉)
University of Portsmouth, Portsmouth, UK
e-mail: Alessandro.melis@port.ac.uk

T. Pievani
University of Padua, Padua, Italy
e-mail: dietelmo.pievani@unipd.it

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_15

components and to both changes of function of tectonic elements and chance of use of the architectural space. In the final part of the chapter, informal architecture will be introduced as a representation of architectural exaptation.



(Graphic credits: Alessandro Melis & Liam Donovan).

The code of this chapter is 01000101 01111000 01100001 01110000 01110100 01100001 01110100 01101001 01101111 01101110.

1 Introduction

Many scholars have focused on transdisciplinary studies involving biology and other disciplines, including glottology, economics, and technology [2, 3].

Despite the potential applications in design practice, there are no significant studies on the relationships between evolutionary biology and architecture if we exclude those on biomimicry.

Beginning with the comparisons, between the phenotypes of the human body, from classical to Renaissance architecture, then codified, for the first time in the primitive hut by Marc-Antoine Laugier [4–6], explorations on biomimicry have today reached high levels of interest, especially in the study of materials technology and bio-inspired technologies such as bio-robotics, and are also promoted by influential educators and scholars such as Neri Oxman [7].

The potential of bio-mimicry lies in the fact that in several cases, we are still far from being able to simulate what evolution by natural selection has produced in billions of years (e.g., photosynthesis). The limit of biomimicry, however, is that any architectural reference to biology stays on a phenotypical level and does not instead consider biology as a potential starting point for a paradigmatic change in the principles of design and its workflow, which, however, in times of global crises caused mainly by the way we design cities, may be necessary [8].

A more recent cross-disciplinary stream of research regards the introduction of autopoiesis in architectural computation design. The term autopoiesis, expressed for the first time by the Chilean Biologist Humberto Maturana, along with his colleague, the late neuroscientist Francisco Varela, is about the self-generating and self-maintaining systems, which influenced system thinking and cybernetics as well as computation architecture.

In this chapter, the interest in the relationship between the biology of evolution and architecture is attributable to the possible analogies between natural selection and architectural design, which could potentially corroborate the relevance of autopoiesis in architecture as an approach aiming at understanding the design process rather than its final product.

This analogy is also confirmed by the several definitions of biologists regarding the principles of evolution and adaptation [9], which, very often, are superimposable with those used by architects regarding systemic architectural design and sustainable projects.

However, despite the relevance of the relationships indicated in the premises, architecture is still considered an autonomous discipline for a large part of architectural theorists, and, generically, refers, in the most well-known history of architecture books, to a quite short timeframe, starting from the beginning of the Bronze Age, up to the present day.

In some cases, what happened before this period is ascribed rather to the field of archaeology, which only selectively is considered relevant for the growth of knowledge in the field of architectural design [10].

Biological and cultural evolution, on the other hand, in terms of only what concerns the relationships between *Homo sapiens* and the habitats, covers a period of at least 200,000 years, since we were born in Africa as a new species.

Also, in this case, the study of ecology shows that a wider time span allows to better measure the effects of global crises, such as climate change, highlighting how, therefore, the study of the biology of evolution offers opportunities for reading the futuristic scenario, more efficiently than those developed over the time of traditional urban studies. Coincidentally, according to Pringle, the extension of the research toward 200,000 years ago, allows us to discover the origin of human creativity and, therefore, its skill to design [11].

In fact, the reading of the evolutionary trends of the city, according to long times (what biologists and geologists call "deep time"), in the architectural perspective, but very short if interpreted in terms of the biology of evolution, exposes the design to errors of interpretations on future trends, precisely on the basis of biology studies. This risk also includes other terms of a possible analogy between biology and architecture, such as the possible reading of the contemporary city in recapitulation form, meaning as such that contemporary cities are the pinnacle of a linear and progressive evolutionary process at the origin of which less advanced urban forms are found.

The limited heteronomy of research in architecture, therefore, in addition to a more limited time span research, has also led to overlooking theories already known instead in the biology, and already considered in other fields of research (archaeology), such as that of "niche construction": the recursive and constructive process by which organisms actively alter environmental states, thereby modifying the conditions that they, and other organisms, experience, and the frame of selective pressures in their environments [12–14]. This kind of selective pressures actively modified by human activities, as an evolutionary account of Anthropocene, has started to be relevant in architecture, quite recently, and mostly through a limited literature review including quite exclusively philosophical trends, such Tim Morton's Dark Ecology [15], rather than hard sciences. It could instead be assumed that the architectural interventions and the shaping of landscapes have long been true drivers of the human niche construction.

The transposition in the history of architecture of criticism of Ernst Heckel's doctrine that ontogeny recapitulates phylogeny [16] could in itself be another significant and original topic for the writing of a chapter.

Given the vastness of the possible analogies between evolutionary biology and architecture, however, the present one is dedicated to an initial study limited to the specificity of "exaptation" or functional shift, which, according to the following paragraphs, can become a key term in reading the cities in terms of resilience.

Consequently, late paleontologist Stephen Jay Gould's seminal work on exaptation will be the main reference of the study.

2 Exaptation Versus Adaptation in Architectural Design

Between the 1960s and 2000s, a relevant discussion took place in the field of biology fueled, above all, by Stephen Jay Gould, regarding the definition of the evolutionary mechanisms of adaptation.

This discussion, which explored a vast existing literature, from Williams [17] to Bock [18–21], resulted in a pivotal article by Gould and Elisabeth Vrba on exaptation [22]. With this text, the authors intended to challenge and extend the traditional taxonomy around the term of adaptation by highlighting that the same referred exclusively to that process through which the form follows a certain function, thus excluding from this process all those cases in which the forms (i.e., pre-existing structures) were subsequently co-opted by an adaptive function.

The two authors come to define aptation as the set of both of these processes (so, any structures or traits someway useful for the fitness), while exaptation concerns functional co-optation [22]. According to Pievani and Serrelli [23], two types of exaptation must be distinguished: 1) exaptation as a functional shift of a structure that already had a function (as in the case of the feathers of birds, which did not evolve for flight, but were already present and connected to thermoregulation and sexual selection functions; the same for fingers and limbs, already present before the ancestors of tetrapods lived permanently on land); 2) exaptation as the use of structures without any function (as in the case of the reuse of dismissed organs or the use of redundant portions of a system). Spandrels refer in particular to the second type.

Exaptation does not exclude adaptation but extends the taxonomy of the processes that contribute to the fitness of organisms and populations. It implies that the current function of a structure (an anatomical or behavioral trait) does not always coincide with its historical origin. The concept was first proposed by Charles Darwin in the sixth edition of *The Origin of species*, to explain the gradual evolution of organs of particular complexity.

Exaptation has so far had interesting applications in the study of the evolution of technologies, where in many cases, a new technology arises from the reuse and re-functionalization of previous tools [24], and the evolution of languages [25], while, in architecture, it has not yet been applied. The potential interest of this discussion by architects regards the fact that, as described by Gould and Vrba [9, 22], the definitions of aptation-adaptation-exaptation can be perfectly transferred in the field of design. Nevertheless, the two disciplinary areas have developed, autonomously, their own linguistic code, from which their respective taxonomies derive. A cross-disciplinary connection is useful when it has a heuristic value, that is, when using a concept from another discipline helps us to ask new fruitful questions in another discipline. It seems to us that exaptation meets this criterion very well, as argued below.

Thus, a transdisciplinary approach could potentially lead to a re-interpretation of the history of architecture through the meaning of function, associated almost exclusively with the rationalism of modern movements [26–30], taking into account the more articulated connotations deriving from biology, which might challenge the conventional architecture positions.

The introduction of exaptation in architecture could be one of these cases. In architecture, in fact all aspects concerning the use of buildings and spaces, not predicted by design, have been excluded from the paradigms of the design processes.

As happened in the field of biology, even for architecture, the absence of a definition does not mean that some forms of functional co-optation have not been detected.

According to Gould and Vrba [22], "taxonomies are not neutral or arbitrary hatracks for a set of unvarying concepts; they reflect (or even create) different theories about the structure of the world. As Michel Foucault has shown in several elegant books (1965 and 1970, for example), when you know why people classify in a certain way, you understand how they think."

As with biology, by questioning the existing taxonomy, they were simply relegated to the realm of random facts, or read as aspects of a sociological or psychological nature, which had nothing to do with the project or design. According to different systems of thought, "What seems peripheral to us becomes central, and distinctions essential to us do not matter" [22].

Similarly, also in architecture "classifications are not passive ordering devices in a world objectively divided into obvious categories. Taxonomies are human decisions imposed upon nature—theories about the causes of nature's order. The chronicle of historical changes in classification provides our finest insight into conceptual revolutions in human thought. Objective nature does exist, but we can converse with her only through the structure of our taxonomic systems" [31].

Publications such as *Architecture without architects* [32], and *Toward a Critical Regionalism* [33] have emphasized alternative architectural paths to design mainstream. And they also proposed a re-interpretation of the link between architecture and archaeology, which corroborated the growing interest in informal and vernacular architecture.

However, the relevance of not foreseen uses in architecture has been considered as a criticism of the current design, often from a sociopolitical perspective, rather than an opportunity to extend the significance of architectural design itself.

According to the taxonomies of architecture, we define it as designing everything that implies a functional determinism: even when this determinism concerns the transformation of pre-existing structures born and developed for different uses. For example, the walls of the early Christian temple which are co-opted to become foundations of a new basilica, or when a house of a celebrity then becomes a museum.

All these transformations modalities, when performed consciously, therefore with determinism, even if the initial architectural design did not imply them, are alternatively indicated as regeneration, restoration, conservation, preservation, and functional adaptation, depending on the geographic context or the historical period, or the specific architectural discipline [34].

However, several forms of functional co-optation in architecture remain excluded from this presumed interpretation of determinism. For example, the phenomena of temporary appropriation of public space, which contributes to the resilience of the neighborhoods, concern the uses of space not planned in any conventional design [35].

According to the studies on the temporary appropriation of public space, the expansion toward unexpected uses, logically, increases the possibility of "survival" ("life cycle" in architecture) of those places, whilst the lengthening of this cycle is an essential condition to reduce the environmental impacts that are at the origin of the current environmental crisis [36].

Hence, the significance of these phenomena is due to the evident similarity with the mechanism of natural selection, which might help in understanding how to radically redefine the paradigms of design in times of climate change and ecological disruption. Natural selection never operates *ex nihilo* and does not produce new structures from zero, but by modifying the existing material, including the historical constraints it contains. So, exaptation shows us that every structure has an intrinsic transformative potential. Exaptation is an extension of the adaptative possibilities of organisms because it underlines flexibility and plasticity, as in the case of the human brain, a compendium of exaptations of neural areas and networks, or in the case of human DNA, 80% of which have no known function and act as a repertoire for possible reuses.

In these phenomena, as in many others, it is a question of giving a name to the ability of a project to allow functions and uses that were not expected, or predictable, before its realization. The projects have inherent potentials and unexpected reuses.

3 Spandrels: The Missing Link Between Architecture and Biology

The most surprising aspect of the lack of transdisciplinary studies linking the biology of evolution and architecture is Gould's use of an architectural metaphor to explain exaptation.

Gould even uses an architectural tectonic component (“spandrel”) to define in biology the characteristic that is a by-product of the evolution of some other characteristic, rather than a direct product of adaptive selection:

> The great central dome of St. Mark’s Cathedral in Venice presents in its mosaic design a detailed iconography expressing the mainstays of Christian faith. Three circles of 2 figures radiate out from a central image of Christ: angels, disciples, and virtues. Each circle is divided into quadrants, even though the dome itself is radially symmetrical in structure. Each quadrant meets one of the four spandrels in the arches below the dome. Spandrels-the tapering triangular spaces formed by the intersection of two rounded arches at right angles are necessary architectural byproducts of mounting a dome on rounded arches. Each spandrel contains a design admirably fitted into its tapering space. An evangelist sits in the upper part flanked by the heavenly cities. Below, a man representing one of the four biblical rivers (Tigris, Euphrates, Indus, and Nile) pours water from a pitcher in the narrowing space below his feet. The design is so elaborate, harmonious, and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense of the surrounding architecture. But this would invert the proper path of analysis. The system begins with an architectural constraint: the necessary four spandrels and their tapering triangular form. They provide a space in which the mosaicists worked; they set the quadripartite symmetry of the dome above. Such architectural constraints abound, and we find them easy to understand because we do not impose our biological biases upon them. Every fan-vaulted ceiling must have a series of open spaces along the midline of the vault, where the sides of the fans intersect between the pillars. Since the spaces must exist, they are often used for ingenious ornamental effect. [1]

Architectural constraints have a structural function, independent of their subsequent artistic and symbolic use. Although the use of the architectural term, spandrel, in biology has turned more than forty years old, the biological term, exaptation, vice versa, is absent in the architectural literature with the exceptions of few very recent papers by Furnari [37, 38] and Faulders [39].

The case studies explored by these authors have shown "that in design, like in biology, innovation-by-exaptation can be usefully contrasted to innovation-by-adaptation, which assumes evolution of the structure of a feature toward better function. In contrast, exaptation describes the unforeseen connection between an existing feature and a new function, different from the function for which the feature was originally designed or selected for" [38].

With exaptation, evolutionary thinking stresses the role of constraints in evolution. A constraint (physical, developmental, structural, and so on) is not just a negative limitation to change, but an opportunity to re-use something already there. Natural selection works in the economy: it is less costly to use a structure already existing, rather than to evolve it *ex nihilo*. This is also true for molecular evolution, when de novo genes evolved from already existing scraps of DNA without any previous function. Translated in architecture, therefore, exaptation will include all the overlooked forms of functionalization which are not design based. In this case, we mean, for design, the deterministic approach to design, which implies a specific use (function) attributed to architecture when in place.

Similar to what has happened in biology, a change in the taxonomy has implied a change in the perspective as well. The indeterministic forms of functionalization, in fact, seem much less collateral to the deterministic ones, once we decide to classify them.

The classification could include at least categories such as: (1) the functionalization of existing geomorphologies; (2) the re-functionalization of existing designed structures; (3) the integration of new functions to a certain architecture; (4) and the change of use.

Each category might consider a new functionalization that was not implied by the initial design or by the existing context. The re-functionalization might take place formally, through a conventional design, or informally.

The re-functionalization through a new set of deterministic designs has acquired sufficient dignity to be acknowledged and classified, mostly, during Romanticism, in which the historical and monumental significance of architecture has become a cultural value.

From the roots of the nineteenth century derive today's terms which span from the traditional "restoration", "conservation", and "preservation", to the more recent "adaptive reuse", "functional transformation", and "regeneration".

Surprisingly, considering that the impact, in terms of inhabitants and users, of informal architecture is equal to the formal one, none of the aforementioned terms includes the correspondent functionalization which takes place without a deterministic design process.

It could be said that, even without an adequate distinction with no nuances, these transformations or functionalizations could fall within a generic idea of informal architecture, according to Table 1.

4 Informal Behavior as a Form of Exaptation

Through the lens of biology studies, this part of the chapter is an initial discussion on the informal hypothesis as a form of exaptation, in opposition to the conventional perspective of informality as a result of urban environment deterioration.

According to Gould [31], taxonomies in biology can be affected by prejudices and may incur errors due to the identification of trends that, in reality, compared to a time period or to a larger population, turn out to be simple fluctuations:

> The common error lies in failing to recognize that apparent trends can be generated as by-products, or side consequences, of expansions and contractions in the amount of variation within a system, and not by anything directly moving anywhere. Average values may, in fact, stay constant within the system (as average batting percentages have done in major-league baseball, and as the bacterial mode has remained for life)—while our (mis) perception of a trend may represent only our myopic focus on rare objects at one extreme in a system's variation (as this periphery expands or contracts). And the reasons for expansion or contraction of a periphery may be very different from causes for a change in average values. Thus, if we mistake the growth or shrinkage of an edge for movement of an entire mass, we may devise a backwards explanation.

Apparently, this error can also be encountered in architecture. In fact, generically, the term "informal," associated with architecture, always has been overlooked if compared with the formal or deterministic perspective, even if, in terms of world population, the impacts of formal and informal settlements are equivalent [40].

There was a tendency to associate the informal exclusively with slums in the past. The United Nations, in fact, defines slums as settlements with inadequate access to safe water, sanitation, and other infrastructure, subject to structurally poor housing quality, overcrowding, and insecure residential status [41].

Today, a wider vision of informal architecture prevails, which also includes positive phenomena of temporary appropriation of public and semi-public space, which, thanks to an increased degree of diversity and inclusiveness, can be considered a measure of social sustainability of a certain settlement and resilience of those communities [35].

And, in fact, it is easy to re-read in some of these phenomena of positive adaptation, not foreseen in the urban design, those same characteristics of plasticity, functional opportunism, and resilience that, in biology, is subsumed in the term exaptation (type 1) and spandrel.

According to Lara-Hernandez, the informal behavioral patterns, which increase the resilience of public spaces, regard mostly commerce, leisure, sport, and worship. He has shown in his research, for example, how an obsolete telephone booth in central Mexico City can become a cooking space to prepare street food or how an

Table 1 This table, realized by the authors, is a first attempt to systematize the phenomena of exaptation in design with the aim of demonstrating the relevance of the non-deterministic approach in architectural design

Deterministic design (adaptation)	Indeterministic design (exaptation)/informal design							
Design based construction	Functionalisation of existing geomorphologies		Re-functionalisation of existing structures		Integration of function		Change of function	
Formal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
I.e. Ville Savoye	I.e. Domus De Janas	Cave	Basilica (pre-existing structures)	Slums (Kenema)	San Marco Spandrels (?)	I.e. Borboletta (Figs. 1, 2 and 3)	Piazza Anfiteatro (Lucca)	Oases Documenta —Kassel

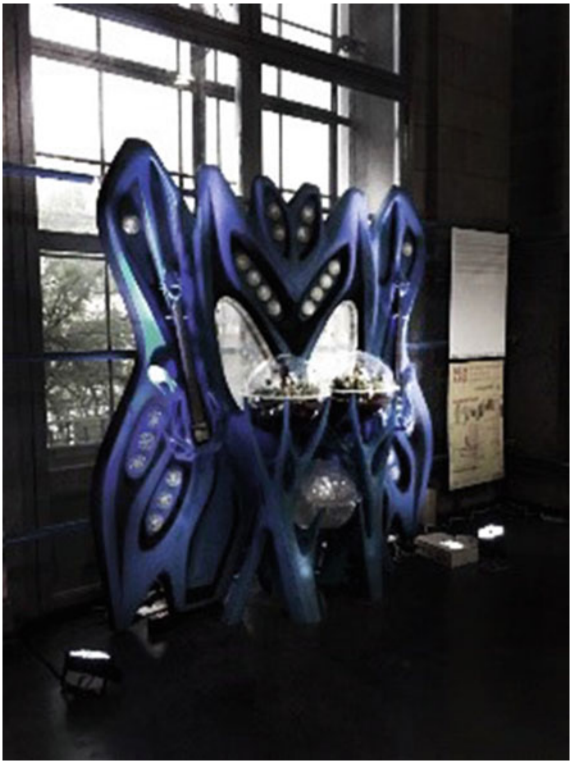


Fig. 1 Borboletta. Bienal Arquitettura. Buenos Aires, 2019 (Photo credit: Monad Studio)

occluded window becomes an altar [35]. Similarly, according to Khemri, a wall of Algiers' El Houma can become either an exhibition space or a place where it is possible to organize a community event within the context of a funeral or a wedding [42].

Despite these and many other informal resilience potentials, the reason for their underestimation remains to be asked and understood.

The prejudice toward informal architecture takes on the tones of mockery if we admit that, today, the first cause of the climate crisis lies in the way in which the formal city has been designed and developed.

In fact, ca. 40% of CO_2 emissions depend on the deterministic design in architecture [40], which, therefore, has shown its limits in overcoming a global crisis, such as the environmental one. Population growth and urbanization are the third major cause (after deforestation and invasive species) of the ongoing mass-extinction of biodiversity.

According to *Science*, over 50% of the world's population lives in urban environments and that implies unintended pressures on global ecology and humans [43]. Human intestinal microbial diversity declines with industrialization and life in urban contexts, giving a dramatic contribution to the rise of many modern diseases.

Fig. 2 Borboletta. Bienal Arquitettura. Buenos Aires, 2019. Borboletta is an architectural practice-based research which simulates the potential of the concept of spandrel in coexistence with non-human species. (Photo credit: Monad Studio)

The World Health Organization states that the impact of climate change on clean air, safe drinking water, and nutritious food supply will cause, between 2030 and 2050, "approximately 250,000 additional deaths per year, from malnutrition, malaria, diarrhea, and heat stress alone. The direct damage costs to health are estimated to be USD 2–4 billion per year by 2030," This will have major effects on informal settlements since "areas with weak health infrastructure—mostly in developing countries—will be the least able to cope without assistance to prepare and respond" [44].

Snyder et al. [45] stated that "these conditions are the perfect breeding ground for EVD (Ebola virus disease). Previous EVD outbreaks occurred in rural and geographically isolated communities. The presumed introduction of the virus to the slums of Kenema and Freetown in Sierra Leone has undoubtedly augmented its spread. Sierra Leone is urbanizing at a rate of 3% each year, and in 2005 more than 97% of its urban population lived in slums." The degradation of the ecosystems in which virus-reservoir animals live adds up, therefore, to urbanization with these characteristics: a lethal mix which, due to human activities, makes pandemic outbreaks much more likely.

Fig. 3 Borboletta. Bienal Arquitettura. Buenos Aires, 2019. Borboletta Research group: Alessandro Melis (Heliopolis 21); Eric Goldemberg & Veronica Zalcberg (Monad Studio); Francesco Lipari (OFL); Dzhumhur Gyokchepanar (University of Portsmouth); Jorge Cereghetti (UADE Labs). (Photo credit: Monad Studio)

All this does not exclusively concern the southern hemisphere and the slums, but the weakest fringes of Western society as well. Diane Yentel, president of the National Low Income Housing Coalition in the USA, reiterated the condition of the particular vulnerability of homeless people [46].

It is a paradox. If, on one hand, the organized city contributes to negative environmental pressure, on the other hand, it is the informal settlement that suffers most of its negative effects, though, sometimes, it is there that you can find creative and unexpected solutions regarding solidarity and special practices, such as the ones described in Mexican, Algerian, and many other forms of community resilience.

Nevertheless, we are usually faced with the unilateral idea that the informal city is the problem.

Once more biology, and Gould himself, may offer an interpretative hypothesis to understand why non-deterministic architecture is considered marginal.

Determinism can, in fact, be read as a manifestation of the creative, and even promethean power, which mankind attributes to its alleged privileged position on the evolutionary scale [31].

Gould's position is corroborated by Freud, according to whom the main revolutions in knowledge have led to the dethronement of human arrogance from Olympus of our cosmic certainties; from the Copernican revolution to the discovery of the unconscious, through Darwin's theory of evolution [47].

As has happened during the revolutions of heliocentrism, Darwinian evolutionism and the discovery of the unconscious in opposition to rationalism, today, it is the environmental crisis that dethrones humanity and brings it back to the margins of nature and, hopefully, within the ecosystem.

> "Much as we may love ourselves, *Homo sapiens* is not representative, or symbolic, of life as a whole. We are not surrogating for arthropods (more than 80 percent of animal species), or exemplars of anything either particular or typical. We are the possessors of one extraordinary evolutionary invention called consciousness—the factor that permits us, rather than any other species, to ruminate about such matters (or, rather, cows ruminate and we cogitate). But how can this invention be viewed as the distillation of life's primary thrust or direction when 80 percent of multicellularity (the phylum Arthropoda) enjoys such evolutionary success and displays no trend toward neurological complexity through time—and when our own neural elaboration may just as well end up destroying us as sparking a move to any other state that we could designate as 'higher'?"

This preliminary consideration by Stephen Jay Gould, in *Full House* [31], leads us to reflect on how much our society has been conceived to respond to a "superior" perception of ourselves, and, therefore, to last for a limited time. The city, conventionally planned, as the most "advanced" product of man's neuronal capacity, more than any of its other products, suffers from the poor resilience of this approach.

The COVID-19 pandemic showed all our vulnerability and the areas with the greatest (formal and informal) conurbations were clearly the most affected, because the virus exploits our sociability and density of movements, turning them against us.

Continuing in the transposition, why, therefore, do we continually portray the pitifully limited image of the human settlement, in the form of a conventional city, village, or other, which, instead, is nothing more than a brief episode in the life of vertebrates, as if it were the more advanced multicellular coexistence model? And why, then, are we fighting an environmental war to keep alive a form of settlement that inevitably seems to lead us to self-destruction?

According to Gould, we are "narrative creatures," and, as such, we seek directionality, a trend toward which to turn, even if this is not real. For these reasons, before building an idea of the city, we must build a new narrative that leads to an idea of humanity as an alternative to the current one, less privileged along a non-existent evolutionary single scale. We are a tiny branch in the great tree of biodiversity, as we dramatically discover each time during a pandemic.

It follows that the new paradigms of planning, in a crisis phase such as the present one, which presupposes a revolution in human thought, will imply a less "arrogant" vision of human settlements (cities?), as more advanced outposts of life on this planet.

The hypothesis of a new paradigm, however, does not imply renunciation of creativity, as a non-promethean position might imply. According to the paleoanthropologist Heather Pringle, in fact, the variability—intended as the proliferation of forms and artefacts without a predefined use—is at the origin of creativity. Diversity, in biology, is the fuel of any change. Art, technology, science, and all the expressions of creativity are a manifestation of associative thinking, activated when linear logic, which is the standard modality of survival, and its specialized and one-dimensional expressions in the urban realm are unable to respond to the crisis [11].

Thus, in the analogy between biology and architecture, the role of non-intentional nature, in the first case, is attributable to the designer, in the second.

Future cities, or rather the forms, today unimaginable of coexistence, will depart from the idea that architecture and nature are separate and equivalent conditions that are compared on a game table, sometimes in harmony, often in conflict.

> The chess-board is the world; the pieces are the phenomena of the universe; the rules of the game are what we call the laws of Nature. The player on the other side is hidden from us. We know that his play is always fair, and patient. But also we know, to our cost, that he never overlooks a mistake, or makes the smallest allowance for ignorance. [48]

This allegory, known as Huxley's chessboard, discussed in biology for some time, is still present, in the 70s in the famous diagram by M Scott Pegg, which accompanies the bestseller *The Road Less Travelled*. Here, we could say that, following the same idea of linear progress, we architects have built cities to this day.

In times of global environmental crisis, of anthropic origin, this apparent diversity of subject seems even more relevant, given that, in ecological terms, the hope is that man will overcome the projection of his role as an alternative to nature, as an entity emancipated from ecosystem dynamics [31].

Similarly, to the mechanisms described by evolutionary biology, the architectural spandrels imply variability of forms and their potential relationships with the environment, which allows nature to adapt to unpredictable environmental conditions. Exaptation means that a degree of structural redundancy is frequently functional, mostly in ever-changing and unstable environments. The most creative systems in nature (genomes, brains, and ecological networks) are all redundant. Conversely, functional specialization can be very risky in times of environmental change.

Therefore, the survival and balance of ecosystems also depends on architectural exaptations intended as an urban redundancy of forms and relationships which favor the adaptation to continuously changing environments.

As well as exaptation in evolutionary biology, informal architecture is an opportunity aimed at the resilience of the city bodies and could lead to redefining the paradigms of human coexistence and the urban fabric in times of global crises.

Without having the arrogance to provide a conclusive answer, the change of perspective raises a series of hypotheses, which, hopefully, will be the subject of further studies in the near future. For instance, what can we learn from informal behavioral patterns in terms of low consumption of energy and resources, and about coexistence with non-human species?

This position is, therefore, the consequence of the failure of the order-oriented approach to planning and the awareness of the end of an alleged heroic phase of exclusively deterministic colonization of the planet.

Thus, we expect that the architectural version of the sixth finger (the so-called panda's thumb) of the panda as well as the dinosaur feathers [23, 49] can be recognized in informal settlements and indeterministic architectures. Exaptation implies that each trait has not only a function, but a range of potential side-effects: exploring them, we open new evolutionary possibilities.

In conclusion, architectural exaptations can be considered a non-pre-organized way to colonize the biosphere, which in biologic terms, could also correspond to a non-anthropocentric way to do this.

5 Conclusions

In the chapter's introduction, we have underlined both the significance and the originality of the study of the exaptation as a possibility to extend the architectural design toward more sustainable approaches aimed at enforcing urban resilience.

Exaptation has been described, in the following paragraph, as a functional shift of a structure that already had a prior but different function. Examples of the aforementioned mechanism have included the feathers of birds, which did not evolve for flight, but were already present and connected to thermoregulation and sexual selection functions.

In architecture, a functional shift of a structure that already had a function may apply to forms of decorations embedded in architectural components, such as pinnacles, columns, capitals, beams, and other structures. The coexistence of narrative, symbolic, and aesthetical features together with structural functions are particularly present in gothic and classical architectures in ways in which it is sometimes difficult to distinguish what the primary intent was.

Although the coexistence of these aspects in architecture is well-known, the reading of its origin in a structural key, in the history of architecture literature, and in a narrative key, in the text of art history, respectively prevail, if compared with more syncretic and articulated perspectives.

As mentioned, the first attempt to systematize this matter was probably due to the tradition of architectural treats which, especially in the eighteenth century, indicate in the imitation of nature the origin of the forms, thus placing both structural and decorative reasons in the background.

At the center of the third chapter's paragraph, a second and complementary definition of exaptation, thus not alternative to the previous one, refers to the use of structures, called spandrels, without any function (as in the case of the reuse of dismissed organs or the use of redundant portions of a system).

This second definition, although more ambiguous in the transposition into architecture, given that the architectural example from which Gould draws inspiration for the term spandrel, could fall more easily in the first definition (spandrels have a previous function, a structural one), nevertheless offers greater hints of current interest. It, in fact, can be reflected in non-deterministic and informal planning as an opportunity to increase the resilience of cities, since non-deterministic design does not exclude deterministic planning, in addition to which exaptation does not exclude adaptation, and, therefore, extends the taxonomy of the processes that give fitness to organisms.

The case of the Piazza Anfiteatro, in Lucca, could well-represent non-deterministic design as an extension of the adaptative possibilities of cities [50].

Just as a thumb of a panda, where the panda is the city, the middle age square grows on the abandoned and obsolete remains of the ancient Roman amphitheater (second century AD), which determined its elliptical closed shape, which perfectly fits for the use of workshops, commercial activities, and compact residential quarters, facing a central public space.

From the evolution of the Piazza Anfiteatro clearly a non-deterministic design approach emerges, one which implies that the current function of a structure does not always coincide with its historical origin, while explaining the complexity of city evolution.

As a consequence of further generative role of constraints, the elliptical space, cleaned of medieval superfetations, became the ideal location for the nineteenth-century market designed by the architect Lorenzo Nottolini, required by the economic and social needs of the rise of the bourgeois city. A constraint became an opportunity.

Although not explicitly referable to the theoretical framework of the exaptation studies, attempts to rethink design paradigms in an adaptive key, through trans-disciplinarity, have not been missing, albeit marginalized by the mainstream of architectural autonomy.

Among these, a pivotal role should be attributed to radical movements (during the Sixties and Seventies in the Twentieth century) and, especially, to the Japanese Metabolists and Austrian Radicals. Although unconsciously, the Oases of Haus Rucker Co., presented for the first time at Kassel's Documenta, represent a surprising insight into the possibilities of using exaptation to design in an ecological key. Fueled by the constraints of the limit of the planet's resources, as described for the first time by the Club of Rome (1972), some of those ideas, now almost half a century old, have achieved acknowledgment only today [51], when the consequences of the environmental crisis have become indisputable.

The last paragraph, before the conclusion, focuses on informal architecture as a representation of architectural exaptation.

Despite its great potential, if we exclude the experiences of practices like Urban Think Tank, Teddy Cruz, Giancarlo Mazzanti, and Alejandro Aravena / Elemental, the study of the informal has often been underestimated by architectural criticism.

In these architectures, and especially in Aravena's Quinta Monroy plan (2004), the flexibility and plasticity potential of non-deterministic design has been emphasized by their authors, although a social and solidarity reading of the informal phenomenon has prevailed over the biological analogy, according to which the city could become a compendium of networks.

In conclusion, the use of exaptation's definition, in architecture, raises a few significant questions regarding the evolution of the cities, which corroborate the heuristic value of the cross-disciplinary studies on biology and architecture, which seems even more relevant in times of climate change and global crises.

Exaptation means that a degree of structural redundancy is frequently functional, mostly in ever-changing and unstable environments. The most creative systems in nature (genomes, brains, and ecological networks) are all redundant. Conversely, functional specialization can be very risky in times of environmental change.

Exaptation implies that each trait has not only a function, but a range of potential side-effects: exploring them, we open up new evolutionary possibilities.

Limiting the design to responses to conventional mottos such as "Form follows function" or "Less is more," can be a risk as well.

Thus, in order for architecture to become an opportunity from a problem, it should respond to design paradigms in which redundancy and variable diversity of structures reflect (and enrich) functionalism.

According to this interpretation, the failure of a planned function, due to physical, developmental, and structural constraints, does not necessarily have to be interpreted negatively for the future of the city. It can also be an opportunity to re-use a structure designed for an obsolete function, to respond to unexpected constraints. Like natural selection, the city should work in terms of economy: it is less costly to use a structure already existing, rather than to build it from scratch.

Planners, therefore, will increasingly have to ask themselves which design tools can guarantee greater redundancy and flexibility in the design of cities, rather than aiming at a set of functions for future urban visions with specialized functions, such as those of the traditional "zoning," which, besides being proven wrong, did not allow conversion of the structures into new more virtuous uses.

If diversity, as in biology, is the fuel of any change, a minimalist perspective, based on homogeneity and standardization, unlike what was assumed by the orthodox rationalism of the Modern Movement positions, could, therefore, appear less functional today than we have ever imagined.

Core Messages

- The most creative systems in nature are redundant.
- Exaptation is a functional shift of a structure that already had a prior but different function.
- Structural redundancy is frequently functional, mostly in ever-changing and unstable environments.
- Both in biology and architecture, the redundancy of structures contributes to the resilience of a system in times of environmental change.
- Architecture should respond to design paradigms in which redundancy and variable diversity of structures reflect functionalism.

- The failure of a planned function of a city can be an opportunity to re-use a structure designed for an obsolete function, to respond to unexpected constraints.

References

1. Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc Royal Soc London. Ser B. Biol Sci 205 (1161):581–598
2. Andriani P, Carignani G (2014) Modular exaptation: a missing link in the synthesis of artificial form. Res Policy 43(9):1608–1620
3. Dew N, Sarasvathy SD, Venkataraman S (2004) The economic implications of exaptation. J Evol Econ 14(1):69–84
4. Buchli V (2020) An anthropology of architecture. Routledge
5. Kaufmann E Jr (1964) Memmo's Lodoli. Art Bull 46(2):159–175
6. Kruft HW (1994) History of architectural theory. Princeton Architectural Press
7. Oxman N (2011) Variable property rapid prototyping: inspired by nature, where form is characterized by heterogeneous compositions, the paper presents a novel approach to layered manufacturing entitled variable property rapid prototyping. Virtual Phys Prototyping 6(1):3–31
8. Melis A (2020) A future on resilience: architecture drawings and interview
9. Pievani T (2008) Exaptation. Il bricolage dell'evoluzione. Bollati Boringhieri
10. Melis A, Davis MJ, Balaara A (2017) The history and invocation of the Arche in Austrian radical architecture thinking. Cogent Soc Sci 3(1):1368366
11. Pringle H (2013) The origins of creativity. Sci Am 308(3):36–43
12. Lewontin RC (1982) Organism and environment. In: Plotkin HC (ed) Learning, development and culture. Wiley, New York, pp 151–170
13. Laland KN, O'Brien MJ (2010) Niche construction theory and archaeology. J Archaeol Method Theory 17(4):303–322
14. Laland K, Matthews B, Feldman MW (2016) An introduction to niche construction theory. Evol Ecol 30(2):191–202
15. Morton T (2016) Dark ecology: for a logic of future coexistence. Columbia University Press
16. Gould SJ (1977) Ontogeny and phylogeny. Harvard University Press
17. Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton, New Jersey
18. Bock W (1967) The use of adaptive characters in avian classification. In: Proceedings of the XIV international ornithological congress. Oxford 24–30 July l966, pp 66–74
19. Bock W (1979) A synthetic explanation of macroevolutionary change-a reductionistic approach. Bull Carnegie Mus Nat Hist 13:20–69
20. Bock WJ (1980) The definition and recognition of biological adaptation. Am Zool 20:217–227
21. Bock WJ, Von Wahlert G (1965) Adaptation and the form-function complex. Evolution 10:269–299
22. Gould SJ, Vrba ES (1982) Exaptation-a missing term in the science of form. Paleobiology, 4–15
23. Pievani T, Serrelli E (2011) Exaptation in human evolution: how to test adaptive vs exaptive evolutionary hypotheses. J Anthropol Sci 89:9–23
24. Arthur WB (2009) The nature of technology: what it is and how it evolves. Simon and Schuster

25. Lass R (1990) How to do things with junk: exaptation in language evolution. J Linguist 26 (1):79–102
26. Benevolo L (1977) History of modern architecture, vol 2. MIT Press
27. Collins P (1998) Changing ideals in modern architecture, 1750–1950. McGill-Queen's Press-MQUP
28. Scully VJ (2003) Modern architecture and other essays. Princeton University Press
29. Tafuri M, Verrecchia G (1980) Theories and history of architecture. Granada, London
30. Zevi B (1978) The modern language of architecture. Australian National University Press
31. Gould SJ (1996) Full house: the spread of excellence from Plato to Darwin. Harvard University Press, Cambridge, MA
32. Rudofsky B (1987) Architecture without architects: a short introduction to non-pedigreed architecture. UNM Press
33. Frampton K (1993) 20 Towards a critical regionalism: six points for an architecture of resistance. Postmodernism: a reader, 268
34. Carbonara G (ed) (1997) Trattato di restauro architettonico. Utet
35. Lara-Hernandez JA, Melis A (2018) Understanding the temporary appropriation in relationship to social sustainability. Sustain Cities Soc 39:366–374
36. Curran MA (ed) (2016) Goal and scope definition in life cycle assessment. Springer, Berlin
37. Furnari S (2009) Mechanisms of aesthetic exaptation in architecture: how a Beauxarts Garden evolved into an avant-garde art park. Paper presented at the XXV EGOS Colloquium, July 2–4, Barcelona, Spain (Winner of the EGOS 2009 Best Dissertation Paper Award)
38. Furnari S (2011) Exaptation and Innovation in architecture: the case of Chicago's millennium park
39. Faulders T (2014) Radical authenticity: exaptation in architecture. In: ACSA/AIK international conference, Seoul, South Korea
40. Melis A, Medas B (2020) Tecnologie avanzate per la resilienza dell'architettura e della comunità. Bioarchitettura: Appunti per una città sostenibile. Nardini Editore, Lucca, IT
41. Warah R (2003) The challenge of slums: global report on human settlements 2003. Retrieved from https://www.globalpolicy.org/component/content/article/211/44579.html
42. Khemri MY, Melis A, Caputo S (2020) Sustaining the liveliness of public spaces in El Houma through placemaking. J Public Space 5(1):129–152
43. Dominguez Bello MG et al (2018) Preserving microbial diversity. Science 362:33–34
44. World Health Organization (2018) Climate change and health. Retrieved from https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health
45. Snyder RE, Marlow MA, Riley LW (2014) Ebola in urban slums: the elephant in the room. Lancet Global Health 2(12):e685
46. Yentel D (2020) The end of the beginning. Retrieved from https://nlihc.org/resource/end-beginning-diane-yentel-nlihc-president-and-ceo
47. Fletcher J (2007) The scenography of trauma: a 'Copernican' reading of Sophocles' Oedipus the King. Textual Pract 21(1):17–41
48. Huxley TH (1868) On the classification and distribution of the Alectoromorphae and Heteromorphae. Royal Institution
49. Gould SJ (1991) Exaptation: a crucial tool for an evolutionary psychology. J Soc Issues 47 (3):43–65
50. Melis A (2019a) Stadiums aren't fated to disrepair and disuse. History shows they can change with the city
51. Melis A (2019b) The introduction of nature in the Austrian radicals practice. In: Planning cities with nature. Springer, Cham, pp 45–63

Alessandro Melis (1969), Ph.D. (Florence) RIBA ARB AOU, is Full Professor of architecture innovation at the University of Portsmouth and the Director of the International Cluster for Sustainable Cities. In 2019, he was appointed by the Italian Minister of Cultural Heritage as the curator of the Italian Pavilion at the Venice Biennale. Previously, at the University of Auckland, he was the head of the technology area and director of postgraduate engagement. In the period 2010–2013 he has been the Director of Brain City Lab at the university of Applied Arts Vienna, guest professor at the Anhalt University Dessau, and an honorary fellow at the Edinburgh School of Architecture. He has also been invited as a keynote speaker at the China Academy of Art, the MoMA New York, the University of Cambridge, TED, and the Italian Institute of Culture in London, and the UNESCO Headquarters in Paris. His work was the object of several publications, including the recent monograph (Rome, 2020) edited by Giuseppe Fallacara, titled "Alessandro Melis, Utopic Real World."

Telmo Pievani (1970) is Full Professor at the Department of Biology, University of Padua, where he covers the first Italian chair of Philosophy of Biological Sciences. He also teaches Bioethics and Science Communication. Since October 2016, he is Rector's Delegate for Institutional Communication, University of Padua. Researcher in the field of Philosophy of Biology, after Ph.D. researches in USA, he has been Professor of Philosophy of Science at the University of Milan Bicocca (2001–2012). From 2017 to 2019, he has been President of the Italian Society of Evolutionary Biology. He is Fellow of several academic Institutions and scientific societies: Istituto Veneto di Scienze, Lettere ed Arti, Class of Sciences, Venice; Turin Academy of Sciences, Class of Biological Sciences; Accademia Olimpica in Vicenza, Class of Sciences and Technique; "Umberto Veronesi" Foundation for the Progress of Sciences, Milan; Istituto Italiano di Antropologia, Steering Board, Rome. He is member of the editorial boards of Evolution: Education and Outreach, Evolutionary Biology, Rendiconti Lincei Sc. Fis. Nat., Istituto della Enciclopedia Italiana Treccani, and Le Scienze, Italian edition of Scientific American. He is author of 256 registered publications.

A New Generation of Shrimp Aquaculture for Sustainability and Resilience of Future Green Cities

16

Carlos R. Rojas-Garcia, Marcos R. Cabezas-Perez, Arturo Ruiz-Luna, and Isabel Jimenez-Garcia

We don't inherit the earth from our ancestors
We borrow it from our children.

Native American proverb

Summary

This chapter focuses on two emerging concepts, green cities and urban farming, as a framework for creating a sustainable and resilient food system. We propose integrating urban shrimp farming in a closed system with microbiome control and total water recycling. We aim to achieve seafood farming in future green cities by extending the notion of what ecological aspects make a city green more resilient and sustainable, including creating local value chains, the local market and non-market activities, and new forms of responsible entrepreneurship. We paid particular attention to creating forms of ecological awareness for the citizens and green urban activism, pursuing the acquisition of ecological notions of the natural shrimp life cycle and protecting the natural mangrove forest that keeps shrimp healthy with a superior genetic pool. Furthermore, we reviewed approaches to democratizing green cities with ecological activism to empower

C. R. Rojas-Garcia (✉) · M. R. Cabezas-Perez
Seafood Sustainable Farming, Research and Consultancy, Mexico City, Mexico
e-mail: rocarlo@protonmail.com

A. Ruiz-Luna
Centro de Investigación en Alimentación Y Desarrollo A. C., Coordinación Regional Mazatlán. Av. Sábalo Cerritos S/N, 82112 Mazatlán, Sinaloa, Mexico

I. Jimenez-Garcia
Tecnológico Nacional de México, Instituto Tecnológico de Boca del Río. Km 12 Carr. Veracruz-Córdoba, Boca del Río, 94290 Veracruz, CP, México

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_16

local-level efforts to coordinate non-corporate relations in food democracy and the goodwill to pay for ecosystem services.

The code of this chapter is 01000101 01100011 01101111 01110011 01111001 01110011 01110100 01100101 01101101 00100000 01110011 01100101 01110010 01110110 01101001 01100011 01100101 01110011.

1 Introduction

1.1 Why Shrimp Farming Suits Future Green Cities?

Sustainable shrimp aquaculture (SSA) is growing beyond the coasts to provide food with high-quality protein produced under low greenhouse gas (GHG) emissions. As any "green venture" it has environmental, social, and economic challenges and rewards, providing local goods and contributing to urban economies, but most important, fostering the appreciation for ecological and aquatic natural systems in town people [1–3]. Together with urban agriculture (UA), a term used by urbanists and urbanites, sustainable shrimp aquaculture is a local effort to produce food within cities. Based on the basket of goods provided by the farm and their marketing toward urbanites, there is a marked division between farms operating for local urban and global markets. SSA in future green cities will not compete with industrial producers and certainly, will promote local food systems and short supply chains using direct marketing techniques such as selling at farmers' markets or directly delivering to restaurants and food cooperatives [1] . Before the recent bloom of shrimp farming in cities, the traditional shrimp sector was almost exclusive of tropical coastal areas that

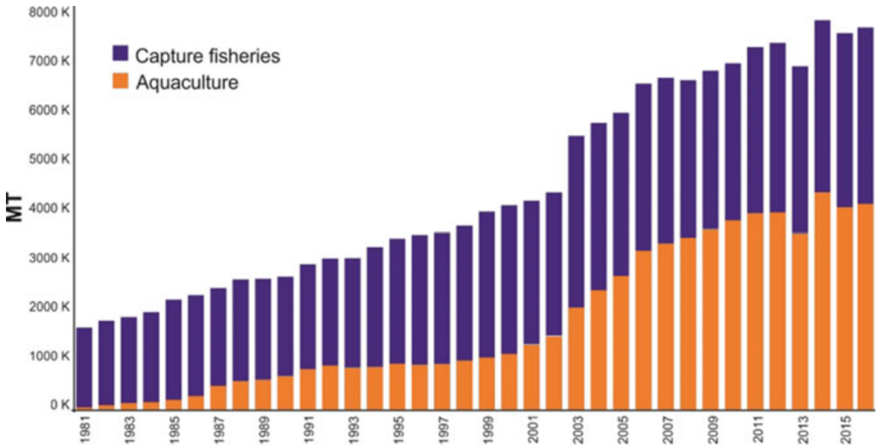


Fig. 1 Shrimp world production, 30% by aquaculture in 2015 [4]

had a rapid expansion, currently accounting for about 30% of total world shrimp production (Fig. 1). Shrimp farming has become highly attractive for investment over the past 30 years, and it is expected that it will play a more critical role in the future due to declining catches of wild shrimp.

Despite its bad reputation and detrimental impact on coastal life and mangroves deforestation, pond farmed shrimp has significant demand from European countries and North America, where shrimp is a commodity, and traditional consumers want to eat it at the lowest price. In the later years, cheap shrimp has contributed to the environmental burden and social inequality.

Mexico and Ecuador are among the ten top biodiverse countries, but it does not make them the most sustainable. These countries have poor aquaculture development compared to Asia, and besides that, they lack appropriate regulatory policies for sustainable aquaculture, favoring the shrimp industry, corporatism, and high-income investors as the primary beneficiaries of the national shrimp resource. Urban or rural small-scale shrimp farming is little known, even socially discriminated against, and inadequately supported by the sector's policymakers. Considering this, the green city movement involves growing our food safety as an alternative to transnational industrial agriculture.

Urban farming is an unstoppable global trend to improve the quality of life and food health. Now the governments and the United Nations (UN) flood us with bureaucratic initiatives for food sovereignty, while the agro-food monopolies continue to poison us, adding soy or refined sugars to everything eatable product. This ugly face of shrimp farming is about to change soon due to several factors such as the "arrival" of disrupting technologies for sustainable shrimp production, driven by the evolution of consumer behavior of new generations, mainly the ecological preferences of young consumers and seafood lovers from wealthy countries, willing "to eat healthy no matter the price"—also by green entrepreneurship in Europe and North America.

Surprisingly, these disrupting technologies were not created in developed countries but came from emerging countries (Mexico and Ecuador). This unusual technological transfer is an emerging route of rich knowledge from the South to the North, equivalent to the pre-Hispanic gold treasures transferred from Latin American native cultures to the European civilization. Shrimp farmers in Latin America just need in return reasonable prices for organic products, work opportunity visas for skilled workers (scarce in EU and North America, the last decade), or education grants. This work focuses on a new generation of aquaculture system developed from scratch in Mexico and Ecuador; we outline the science principles and show a model for urban shrimp farming: the shrimp microbiome tech (SMT) for future green cities.

Even when still prevails the predatory economic premise for "profit no matter the damages," mainly benefiting agro-technology consortiums, as it happens in the Ecuadorian shrimp industry, we bet for an agro-ecological premise for sustainability not matter the cost, bearing in mind the environmental commitment of a new generation of shrimp farmers, only possible in a green city context. Shrimp farming in future green cities must be a good promoter of gender equality and, therefore, the universal right of poor families in emerging countries.

1.2 Green Shrimp Farming: Bioengineering the Shrimp-Water Microbiome

Aquaculture systems can be classified according to the water exchange as open, semi-closed, and closed systems. In open earthen ponds, the soil microbiota cannot be controlled either manipulated. Bacterial changes and microalgal succession are frequent during shrimp farming, often leading to bad water quality, diseases, and mass mortality. However, in closed aquaculture systems, the water microbiota is relatively stable, and it changes according to the nutrient inputs derived from feeding and excretory metabolism. In closed farming without water exchange, the ammonia accumulates rapidly, but this pollution problem has been solved by manipulation of the phytoplankton R and S ecological strategies [5], as well by controlling the water C:N ratio (15:1) to favor the development of heterotrophic bacteria that consume rapidly ammonia (e.g., assuming NH_4–N+NO_2–N+NO_3–N=0.2 mg N/L then requires 3 mg C/L as sucrose to keep C:N ratio = 15).

According to the theory of r-/K-selection, shrimp farmers can prevent diseases from opportunistic pathogens by supplementing the bacterioplankton communities in rearing water with harmless K-strategists [5–7]. Up to recent times, the microbial issues of shrimp farming focused on the elimination of pathogens; however, the classical strategy of cleansing-disinfection-prophylaxis, and even addition of antibiotics, has caused more problems than solutions, due the ignorance about the interactions between the microbes and their environment [8, 9].

A traditional open aquaculture system will typically be characterized by high and unstable nutrient loads, low water retention times, and would be expected to select for opportunistic species type r-strategists [10, 11]. In accordance with the r-/K-selection theory, r-strategists are quick growing opportunists that thrive in niches with low competition and high levels of nutrients; K-strategists, on the contrary, have a slow growth rate, but are efficient competitors for nutrients with high substrate affinity and should produce stable communities at biomass levels, on the brink of the carrying capability. Communities following K-selection are stable to perturbations and with a high diversity compared to communities established below r-selection [10]. Handling of r- and K-populations in recirculation aquaculture systems (RAS) has increased the interest in these practices during later years [12–14], mainly driven by ecological aspects like energy and water reduction and wastes management; the truth is that heterotrophic-based closed systems have been demonstrated to be highly stable and biosecure in the production of shrimp [15]. The recycling of nutrients through the microbiotic food chain rises the food protein utilization efficiency by the shrimp, making feed protein levels decrease [16].

1.3 A New Generation of Shrimp Aquaculture Systems

Our SMT is a new generation system for (highly) intensive farming with the unique particularity that does not rely on higher feed inputs as another intensive aquaculture type [17], but more importantly, its focus is on sustainable approaches, food biosecurity, and saving water. We have achieved the control of particularly tailored (co-cultured) bacteria-microalgae consortia created with freshwater and marine microalgae and beneficial microbes (BM) genus *Bacillus*, *Lactobacillus*, and the yeast genus *Saccharomyces* are ubiquitous to nature: BM is "live micro-organisms that, when administered in adequate amounts, confer a health benefit on the host" [18]. Thus, the microbiotic cycles (nitrogen, phosphorus, and carbon, C:P:N) can be manipulated according to the C:N:P inputs and by-products of shrimp growth, feeding, and excretory metabolism. The first cycle to develop is nitrification, followed by organic carbon and then phosphorous.

Our claims for a natural (ecological) approach of shrimp farming for green cities are based on the fulfillment of the natural aspects of the life history of shrimp in estuarine environments. Once larval development is completed in the open sea (Fig. 2), millions of shrimp postlarvae (PLs) migrate to inland, estuarine waters, usually shallow waters with highly variable environmental conditions (salinity, temperature, and oxygen variations). Mortality under such circumstances is high, reducing the recruitment to around 50% of the total cohort, but survival improves when PLs settle under the protection of mangroves and other wetlands, which also favors food availability based on the detritus food web and provides shelter against predators and high temperatures when shrimp burrows inside the mud during daytime as protective behavior (Fig. 3).

According to natural life processes, our SMT model is totally in accord with shrimp welfare and improves natural living conditions of reared postlarval shrimp. Using shallow rearing water, the feed use for growth is guaranteed, the temperature is controlled, there is no need to look for shelter or burrow, and finally, predation is removed from the system. Constant aeration is provided to supply

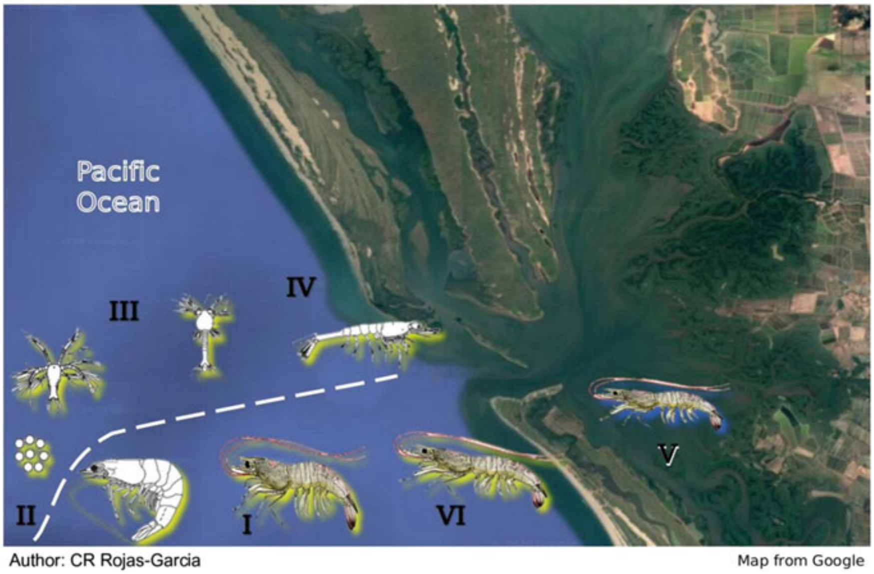


Fig. 2 Schematic Shrimp Life Cycle at Altata Estuary Sinaloa (I—adults growth and reproduction, II—eggs hatching, III—nauplii and metanauplii larval stages, IV—postlarval migration, V—juveniles growth, VI—sub-adults emigration to open sea. *The dotted line* divides the bentic and pelagic stages, the estuary stage V)

oxygen to a high animal density into the rearing tanks, but high density is not animal mistreatment, neither due to the gregarious behavior of shrimp, allowing high-density populations. During farming, the water is also enriched with beneficial microbes and microalgae; therefore, the classical concept of intensive water use for exchange does not apply. The energy CO_2 needed to form one kilogram of edible protein is lower in intensive closed systems than in large ponds. However, most significant is the undeniable fact that manufacturing shrimp for native consumption eradicates massive energy inputs for gathering and transportation. Finally, the nutrients removal by biological filtration as denitrification to eliminate NO_3 or the other nutrients removal represents a waste of resources. Thus, a sustainable approach needs the combination of vegetable crops to assimilate NO_3 and PO_4.

1.4 The Nitrogen Cycle - the Challenge for Nitrates Accumulation

The microbiotic N cycle is composed of several biochemical steps that depend on oxygen. In the aerobic phase (nitrification), dissolved ammonia (NH_4) is converted to nitrate (NO_3) by the bacteria genus *Nitrospyra* by means of two enzymatic steps. In the anaerobic phase (denitrification) occurs the opposite, the accumulated NO_3 is converted to N_2 gas by the anammox bacterial complex (Fig. 4). According to [19],

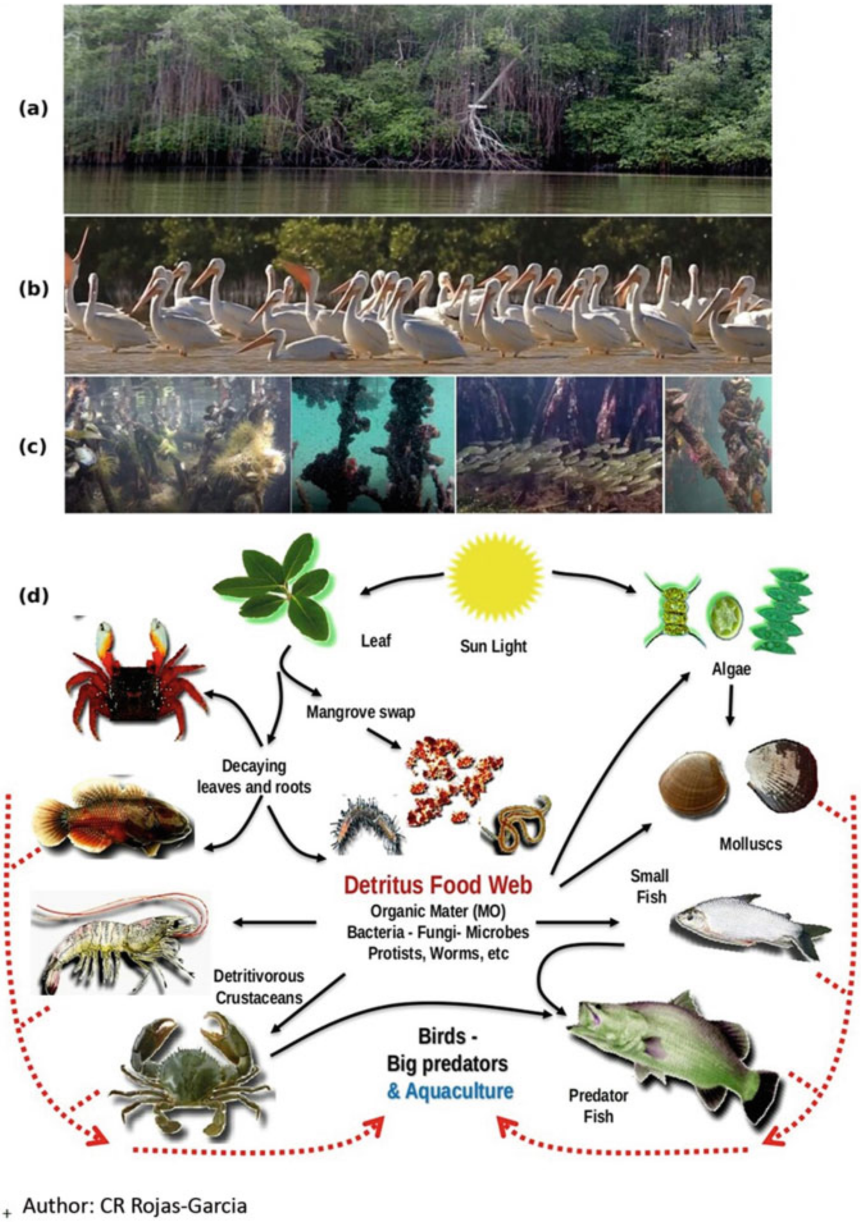


Fig. 3 Mangrove ecosystem and the detritus food web

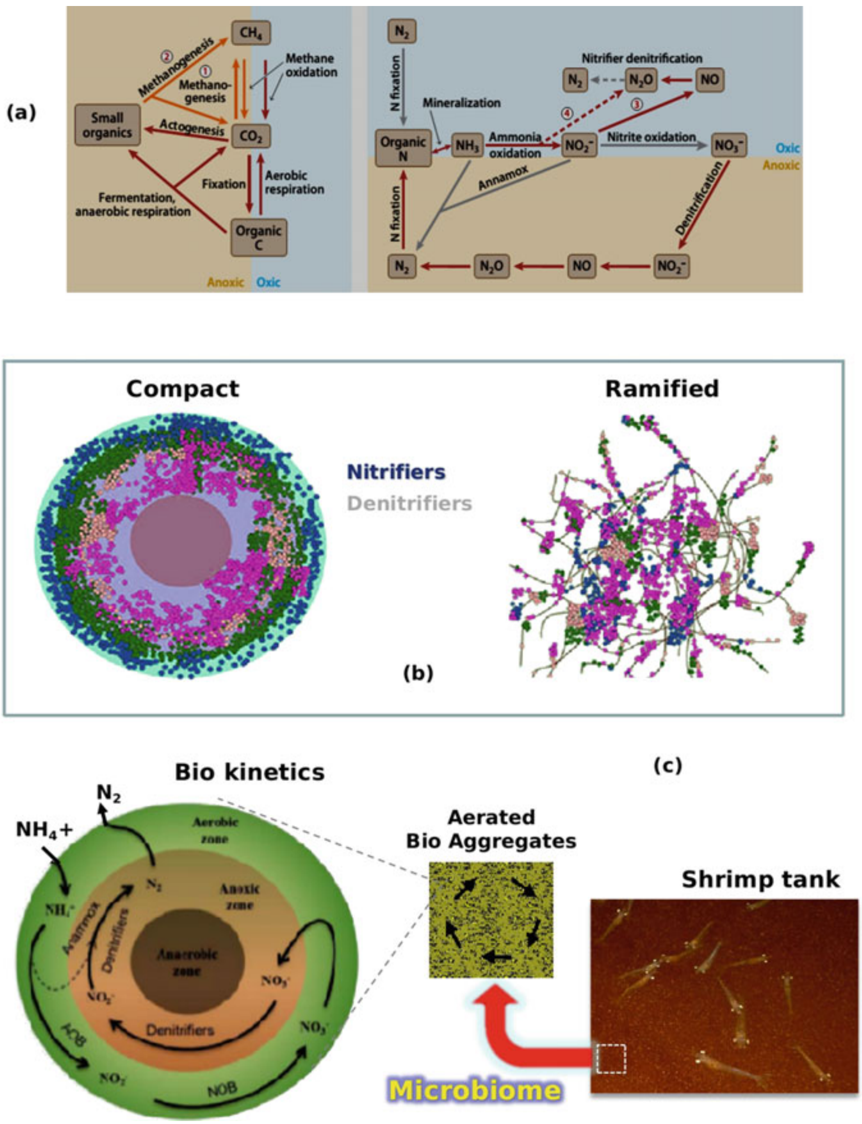


Fig. 4 Microbial ecology [26, 27]: The biofilms microbiome; aggregates, structure, and bio kinetics function

nitritation promotes the activity of ammonia-oxidizing microorganisms and however reduces the activity of nitrite-oxidizing microorganisms. Promoting oxygen-limited, temperature-elevated, or/and pH-raised conditions is common operative way that affords nitritation in reactors [20]. The nitrite-oxidizing step in waste-water treatment systems is performed by *Nitrobacter* (α-Proteobacteria), *Nitrospira*

(Nitrospirae), and *Nitrotoga* (β-Proteobacteria) [21, 22]. Comparing *Nitrobacter* and *Nitrospira, Nitrobacter* have a lower affinity to nitrite than *Nitrospira*; thus, *Nitrobacter* are commonly found in high-nitrite environments, while *Nitrospira* can be found in environments low in nitrite concentrations [23]. In recent years, a whole ammonia oxidizer (comammox) has been discovered in aquarium systems that convert ammonia to nitrate in a very single microbe *Nistropira* [24, 25]. So far, the contribution of comammox to ammonia and chemical group oxidization in nitritation reactors remains to be described. Up to date, NO_3 accumulation is still a significant health issue for farmed species in RAS. NO_3 removal from rearing tanks should be a desired condition, and the denitrification appears to be the apparent microbiotic process; however, it also represents a non-ecological approach since the reduction of NO_3 to N_2O (GHG), and N_2 represents a waste of nutrients as plant fertilizer. The best sustainable approach will be the assimilation of NO_3 into plant biomass; we have tested soilless cultivation of wheat or rice in laboratory conditions.

1.5 Microbiotic Granules for Closed Shrimp Rearing: Ontogeny, Formation Mechanisms, and Functions for Water Amendment

The granules of shrimp microbiome tech are not the same as the sludge granules for wastewater sanitation but similar to periphytic aggregates (Fig. 5). The highly toxic unionized NH_4 must be removed from the system for the well-being of shrimp. In the limitation of water exchange, the alternatives to removing ammonia are microbial and photosynthetic pathways: (1) nitrification by autotrophic bacteria; (2) assimilation by heterotrophic bacteria; and (3) assimilation by photosynthetic algae [28, 29]. In intensive closed farming, oxygen must be supplied constantly to assist the animal growth requirement and the oxidation of toxic compounds, mainly by nitrification and organic carbon cycles. The operative principle to maintain a proper function of the alga-microbe consortium is suspended solids good removal to keep a proper surface area for nitrifying bacteria, the C:N ratios of the microbial feed, the alkalinity of the water, the oxygen near to saturation, and light intensity. Without a proper manipulation of the heterotrophic water, hyper-eutrophication may promote lethal ammonia toxicity. This can be prevented and controlled by the addition of carbon substrates that enhance heterotrophic growth, such as molasses, ethanol, or glycerol [30], and the C:N ratio must be at least 15:1. Additionally, by adding rapid-growing ferments such as *Bacillus*, *Lactobacillus*, and *Saccharomyces* yeast, the growth of the pathogens is inhibited by competitive exclusion. The microbial aggregates formed in heterotrophic systems are microrefineries performing all sets of microbiotic recycling, N, P, and C, resembling a complex engineering process for water sanitation [31]. The nitrification step of the classical nitrogen removal with typically activated sludge methods needs big reactor volumes to achieve the long solid retention time required to allow the slow growth of nitrifying microorganisms. The aerobic granular sludge (AGS) system in water sanitation has gained increasing interest, thanks to its benefits compared to standard activated sludge. Combining AGS with sequencing batch reactors shows that aggregates are relatively easy to obtain, and these had better removal efficiency for organic material and nitrogen removal [31–33], and phosphorus. Filamentous fungi are common components for the granular structure conformation, increasing the surface where the bacteria can attach [34]. Granulation has also been described because of a dynamic floc/particle aggregation and rupture [35, 36]; however, the formation of granules consists of complex steps and is influenced by different physical, chemical, and cellular mechanisms. The AGS reactor conditions enhance the aggregation of microorganisms, shaping the early granules selecting for regular, round, dense, and compact aggregates [37], as shown in Fig. 5. Microbial predation is one of the foremost interactions between living organisms, and a significant explanation for microbes mortality and an important factor in the microbiotic recycling of *Nitrospira* sp. affecting the nitrification performance [38, 39] and Vorticella-like microorganism grazing activity on bacteria in aerobic granules. Finally, the foremost vital modulator of the microbial consortia is the sort of species (filter-feeder, non-filter, and grazers) under culture, as they form the morphology of the aggregates by their feeding habits, like the compact aggregates created by some filter-feeders after ingestion and excretion in fish feces, whereas non-infiltrators could result in completely different forms, from spherical to ramified aggregates. Spherical compact aggregates have an associate anaerobic core with denitrification (anammox) function and external aerobic surfaces where nitrification and phosphorus removal occur. Ramified aggregates have the highest capacity for aerobic recycling of N and P (Figs. 4 and 5).

1.6 Tailoring the Algae-Microbes Consortia

RAS systems can also be classified according to their watercolor as clear, green, and brown water, depending on the living microorganisms within the water (Fig. 5). Green water offers an ecological model to identify beneficial algae-bacteria consortia with the potential to optimize and develop microbial management strategies. According to the r-/K-selection theory, modern shrimp farmers can prevent diseases from opportunistic pathogens by supplementing the algae-microbes consortia with beneficial K-strategists [40, 5, 41 and 42]. Rearing water based on autotrophic chlorophyte community is stable and beneficial for the health and rapid growth of shrimp when the microalgae form a consortium with generalist microbes such as *Bacillus*, *Lactobacillus*, or *Saccharomyces* yeast [43, 44]. Green microalgae manufacture extracellular free fatty acids that most likely attract the microorganism or defend against grazers [45]. So, adding beneficial microbes maintains a stable water microbiota for farmed shrimp's standard growth, health, and well-being.

In contrast, grazing on microbial aggregates by shrimp improves their appetite and digestive functions, stimulating the excretion of organic processing enzymes and maintaining a healthy gut microbiota, likewise as survival and growth [46]. The

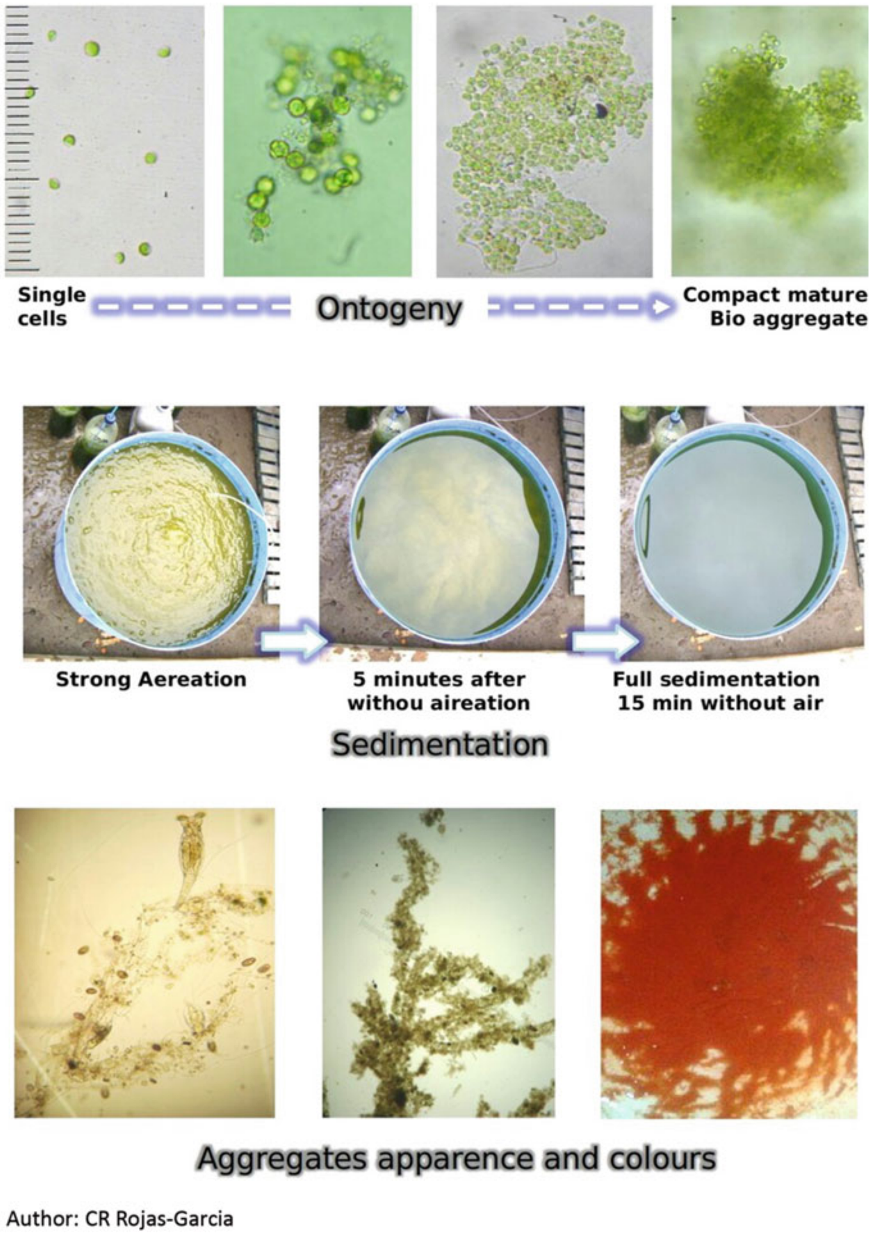


Fig. 5 Ontogeny of the alga-microbe consortium; formation of bio aggregates, sedimentation, and mature forms

dietary inclusion of *Lactobacillus* sp. at concentrations of up to 108 CFU per g increases the activity of various digestive enzymes (protease and amylase) of Pacific white shrimp, and it has been suggested that the combination of *Lactobacillus* sp., *B. subtilis*, and *S. cerevisiae* significantly improves growth performance and survival of shrimp [47–49]. Genus *Bacillus* is a most frequently used probiotic in aquaculture due to its capacity for bacteriocins production, affecting growth performance, strengthening the immune system, and disease resistance against pathogens. A natural resource for screening new quorum quenching bacteria is *Bacillus* species that are commonly regarded safe for use in aquaculture to improve water quality and for disease control purposes [50].

Several green microalgae co-cultured with *Vibrio* bacterium inhibited bacterial growth [51]; additionally, the microalgae remove nitrate and phosphorus from water, positively affecting juvenile shrimp. Also, microalgae are rich sources of carotenoids and accumulate omega 3, 6 PUFAs lipids under nutrient manipulation and mixotrophic cultivation. When shrimp feed on detrital aggregates, adding microalgae-bacteria consortia during nursery growth enriches the natural diet supplies (protein, carbohydrates, lipids, and vitamins) when shrimp feed on detrital aggregates (Fig. 3). The microbiotic function of microbial aggregates in closed systems for shrimp resembles the aerated granular sludge water sanitation but differs markedly in operational parameters, polluting compounds levels, and absence of pathogens.

1.7 Shrimp Microbiome Tech (SMT) Approach for Urban Shrimp Farming

During decades of research and fieldwork, mangrove visits and shrimp farming experience have made it possible to understand some complex biological principles for shrimp rearing in closed systems [52–56]. Thereafter, the translation of these scientific principles and their adaptation to closed urban systems allows describing a simple recipe that common people can use for its benefit. However, due to over-regulative impediments as well as fishermen societies and traditional farmers' inertia, it has been impossible to start up urban–rural shrimp entrepreneurism in Latin American countries, with Mexico and Ecuador, as the main producers. This technology should not be lost in a laboratory archive; instead, it represents a powerful tool offered to next generations living in green cities, contributing to eliminating bad practices carried by industrial farmers and big feed companies that are still causing serious damages to the world ecosystems. The success of urban farmers will demonstrate that the sustainable farming of shrimp is possible anywhere in a clean way. By embracing the sustainability of shrimp farming, future green farmers will be empowered with better knowledge and technical tools to participate in the global decision-making to care for the natural shrimp resources and keep the genetic pool superior, no matter the mangrove is far away from their urban locations. Green farmers will contribute to innovative approaches and new farming experiences. In the next years will see, those old scientific societies, chambers of shrimp trade will not be the sole ones able to organize technical events and international workshops for the exchange of experiences. At international conferences, shrimp experts are scientists or food vendors with no field experience, unaware of daily problems or improvements to lead shrimp to sustainability.

1.8 Basic Components and Steps

The rearing facilities. The tanks or rearing units must be low-cost commercial, garden pools or built with recycled materials, preferentially PET-wood (sheet and bars) or cheap local materials (wood, plywood, river rocks, etc.). The farm operations will require reliable energy for pumping, aerators, and other equipment, which certainly is not a problem in urban areas; however, it still be necessary to have emergency power sources, regarding that reduction in dissolved oxygen levels can kill a crop in a matter of minutes, and access to power emergency equipment will be as necessary for a city as it is in rural areas. As with any new or modified facility, power requirements should be verified against existing grid infrastructure to ensure that loads will not be excessive. Alternative energy sources such as solar, biogas, and hydrogen are preferred for resilience issues.

Step1: Ponds, tanks, and water disinfection. The rearing units must be disinfected prior to adding water, avoiding when possible the use of chloride, preferring other options such as hot water, peroxide, or NH_3 for cleaning surfaces. Access to municipal water eliminates the need for water disinfection, but some salinity adjustment will be necessary. In case of any other water (surface or wells) source, then it is recommended to use hydrogen peroxide for disinfection.

Water use can be scaled up or down according to the size of your tanks based in 1 m^3. Preferably start-up with 4% ppt. For nursery level, 0.2 m water equals 200 L, adding 200 L, raises the level to 0.4 m, and salinity drops to 2% ppt; at week-3, add 200 L with salinity at 1.5 ppt. Certified water quality ranges for shrimp farming can be found in [57].

Step2: Promoting the microbial aggregates in situ. Fill up the water level to 10 cm and then add biofertilizer and microalgae to promote heavy growth and create an algal bloom in the rearing units. The microbial inoculum must be added two days prior to the shrimp stocking. The microalgae-microbe aggregates will create enough food to sustain the postlarval growth during 2–4 weeks.

Step3: Nursery and stocking. Shrimp PL12 acclimated at 4% ppt are stocked to a density of 2000 shrimp/m^2. For the first five days, feeding must be intensive, offering 6–8 ratios of larval feed; following with a commercial diet of 45%P adding protein supplements as a mix of pea + cereals + bovine serum silage and molasses during 25 days, the feed conversion rate (FCR) must be kept below 0.25 (Fig. 6). The last period days until day 60 use high carbohydrate feed adding microbial inoculum and molasses. The FCR must be around 0.5.

Step4: Air supply. The aeration capacity must be 1 HP $\times$ 1-ton shrimp biomass. From day 1 to day 5, aeration is minimum. After this and up to 25 days, increase

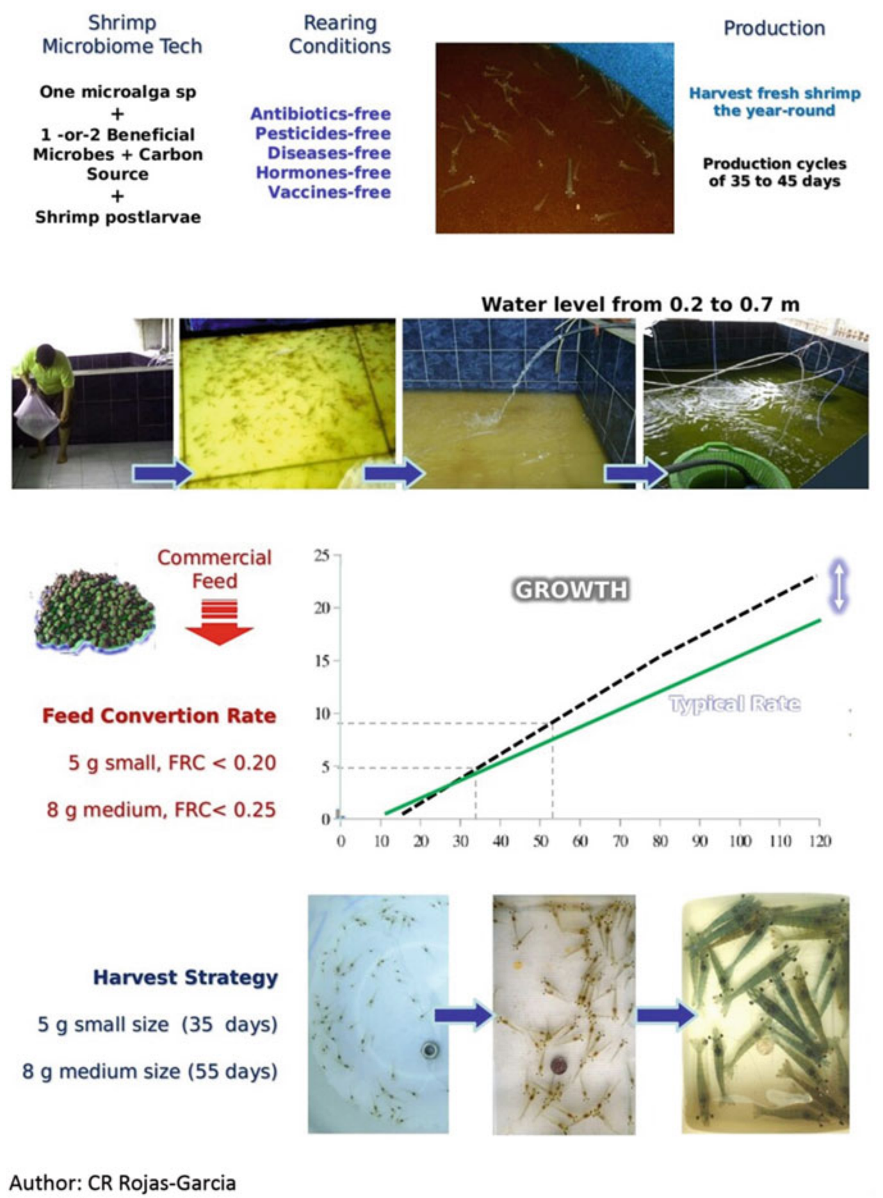


Fig. 6 Shrimp Microbiome Tech (concepts and basic management steps)

the aeration to half-full capacity. For the rest of farming until day 60, apply full aeration day and night.

Step5: Labor, management, and harvesting. Several tasks form the daily routine in the following sequence which are as follows: (1) water quality assessment,

(2) brushing the tank bottoms, (3) meal-1, (4) check feeding trays, (5) meal-2 to meal-6, and (6) check feeding trays 2–6. Growth and population checks every week. A partial harvest or split the population at day 35. Total harvest starts at day 45 over one- or two-week period.

Step6: Harvesting fresh shrimp the round year—consumers benefits. So far, the annual world production of *L. vannamei* has reached 300 million metric tons, half of which is conducted in marine water and the other half in brackish inland water [58]; the future consumption trends for aquaculture products and seafood will be driven by preferences of young consumers and seafood lovers willing "to eat healthy no matter the price." However, these healthy aquatic products must be antibiotics-free, hormones-free, pollution-free, and diseases-free, and these conditions are only possible using the SMT model (Fig. 6). After decades of ambiguity about labeling organic or natural shrimp, it now can be defined and certified locally as the natural labeling, not only in the marketing aspect but also in best farming practices, animal welfare and ecological concerns. Consumers can be certain that they will not eat seafood from banned practices (those devastate marine nurseries). It also will bring new trading, marketing, and economic aspects. Fresh all-year shrimp by city farming; does not need ISC international shrimp certification (being certificated does not mean being sustainable). Big sizes for luxury markets and consumers must be paid green taxes and small sizes more ecological better prices to support local value chains. Small size tastes as good as big size and highly nourishing, but most important is that it takes four times fewer resources to make them.

Food certification makes no sense for locally produced food in green cities, and there is no excuse to attempt any sort of future regulations thriving in urban farm practices that directly market their shrimp to customers, somewhat similar to how a bakery functions. The acceptance and local attitudes for farmed shrimp should be so simple that if the shrimp are raised with good care and tastes good regardless of the size, then it does not need labeling, certification, or official approval. The product might adhere to local acceptance through ancestral channels, *Vox Populi*, or social networks. In the case of urban shrimp, consumers will learn that small sizes are just as tasty and healthy at a lower ecological cost compared to those luxury sizes manufactured at higher ecological costs.

1.9 Biosecurity and Welfare for Shrimp Farming in Green Cities

Basic biosecurity measures are: healthy acclimated postlarvae; water treatment and disinfection; disinfection of culture facilities; optimum water quality management; use of sustainable feeds and proper feeding management; shrimp health monitoring; staff hygiene; control of visitors; and animal pathogen carries.

The biosecurity concept was created from poultry farming experience; on the other hand, the conservative shrimp farmers have neglected health management practices since the moment they were conceived. FAO [59] defines biosecurity as "sets of practices that will reduce the probability of a pathogen introduction and its subsequent spread from one place to another" Safe protocols are intended to maintain the "biosecurity" of an aquaculture facility with respect to certain disease-causing organisms that may not already be present in a particular system. For biosecure shrimp farming, it is required a perfect understanding of the physiology and the environmental parameters of the production system, accompanied by the effective implementation of biosecurity protocols requiring awareness, discipline, and commitment by staff and managers.

1.9.1 Sustainable Feeds and Microbiome Supplements

Shrimp feeds must be designed to meet the specific nutritional requirements; fish meal and fish oil traditionally have been used to fulfill the needs. The sustainability premise requires the incorporation of local feed ingredients and the elimination of wild catch sources or the introduction of genetically modified organisms (GMO). The soybean meal (SBM) has been used to replace the expensive fish meal; the use of SBM in fish diets damages the gut health-promoting alterations in the intestine epithelium, already referred to as SBM-induced enteropathy [60–62]. Most plant feedstuffs have imbalances in nutritional values or lack of essential amino acids such as lysine and methionine and adding crystalline amino acids [63], maybe the solution, but it still is an issue to be solved due to leaching from the pellets during the shrimp feeding activity. Plant protein substitution is limited due to antinutrients and poor palatability; the antinutrients types are protease inhibitors, lectins, tannins, phytates, antigenic or estrogenic factors, and oligosaccharides [64]. Still, the use of plants as a replacement for fish meal is the desired aim for sustainable aquaculture [65, 66]. That can be solved by isolation of protein core separately of antinutrients components [54, 55, 67]. Local feedstuffs of Northern Vietnam like green leave materials (cassava, maize, banana, and grass), grains (rice, maize), fresh sweet potato leaves, cassava root, and manure from numerous sources are available for local feed manufacture.

On the other hand, numerous green microalgae, including *Clorella vulgaris*, *Nannochloropsis* sp., and *Tetraselmis* sp., are have been cultured industrially with expeditiously use of ammonia and phosphorous as nutrients and yielding affordable amounts of the biomass (6.1 g/L). The protein levels can be as high as 60% DW, good amino acids and carotenoid compositions, additionally long-chain unsaturated fatty acids (LC PUFA), and bioactive compounds like natural inhibitors (tocopherol and anthoxanthin). The high level (60%) of algal meal substitution has been tested in fish displaying growth performance and not hampered at any dietary level [68]. The digestibility of *Chlorella* meal by other animals ranged from 68 to 80% [69, 70], and our work also demonstrated assimilation and digestion of algal protein in early larval fish and shrimp [54, 71, 56 and 72]. Beneficial microbes (probiotics) are used to minimize the damage from chronic stress on the fish organism and to increase adaptive plasticity by enriching it with enzymes, essential amino acids, and vitamins. The purpose of probiotics is to heal the intestinal microbiocenosis and increase immunity; in the gut shrimp, probiotic bacteria produce biologically active substances that stimulate the immune system, increase feed conversion, and growth rate [73, 74]. The used sporothermin probiotic is based on the spore forms of bacterial strains *Bacillus subtilis and B. licheniformis.* The use of spore forms allows microorganisms to survive under the influence of stomach acidic environment and under bile action. The indicator of the live bacteria content in the preparation is CFU—the number of colonies forming units. In the sporotherm, the content of *B. subtilis* and *B. licheniformis* is no less than 510 CFU/g [75]. To this far, algae-based shrimp feeds, combined with cereals and legume silage, are the most promising available sustainable feed for shrimp farming of future green cities.

1.9.2 Shrimp Diseases Must not Be a Concern in Urban Shrimp Farming

The biosecurity norms and the good practice manuals created *in-silico* by administrative officials or academics not necessarily suit the urban farmers' real context and have none practical utility. These bad official practices have been a barrier to Latin American countries for aquaculture sustainability, including Mexico where bureaucracy has rooted the 32 states with aquatic health issues, but nothing has promoted aquaculture development. But not only in LA countries the corruption plays a role, today the Norwegian salmon industry is struggling to set up high traditional farming standards for its monopoly benefit attempting to make more complicated the advance of sustainable salmon farming in closed systems [76–79]. There are valid claims from small-scale farmers concerning the inoperative local and international biosecurity regulations and good practices for shrimp farmers drawn up by bureaucrats or academics without proper and updated aquaculture expertise. It is not enough having theoretical expertise when locally is most important the on-farm know-how.

1.9.3 The Eradication of Health Syndromes and Diseases

The use of probiotics stimulates the growth of normal intestinal microbiota, suppressing pathogenic growth [79, 80], and improves the digestibility and the assimilation of nutrients. Additionally, feed costs decrease due to a complete breakdown of food components and the production of vitamins by microorganisms [80]. Once the viral threat became permanent to the shrimp industry, expensive diagnostic techniques were developed, particularly PCR and RT-PCR; however, in the praxis, these solutions were useless for the uneducated farmers' practices and the willingness to pay for the expensive cure, vaccines, and diagnosis, but to pay nothing for systems improvements, animal welfare, and biosecurity. It was until the occurrence of the Early Mortality Syndrome (EMS) that the old industry was obliged to invest in biosecure closed systems where our SMT model (Fig. 6) can play a key role in preventing pathogenic agents and virus attacks [81]. The farmer's awareness of biosecurity increased because EMS attacked the very heart of the industry, the shrimp-seeds.

1.9.4 The Collapse of Shrimp Industry by Viral Diseases

Shrimp farming has suffered catastrophic disease outbreaks that collapsed the global industry [82, 83]. The virus pandemics have served as a "wake up" call that shocked the old industry in Asia and the Americas. From Taura virus (1995) to the White Spot Virus pandemia (2009), the shrimp industry has continued acting negligently to cope with viral and diseases threats [81, 83] (Fig. 7).

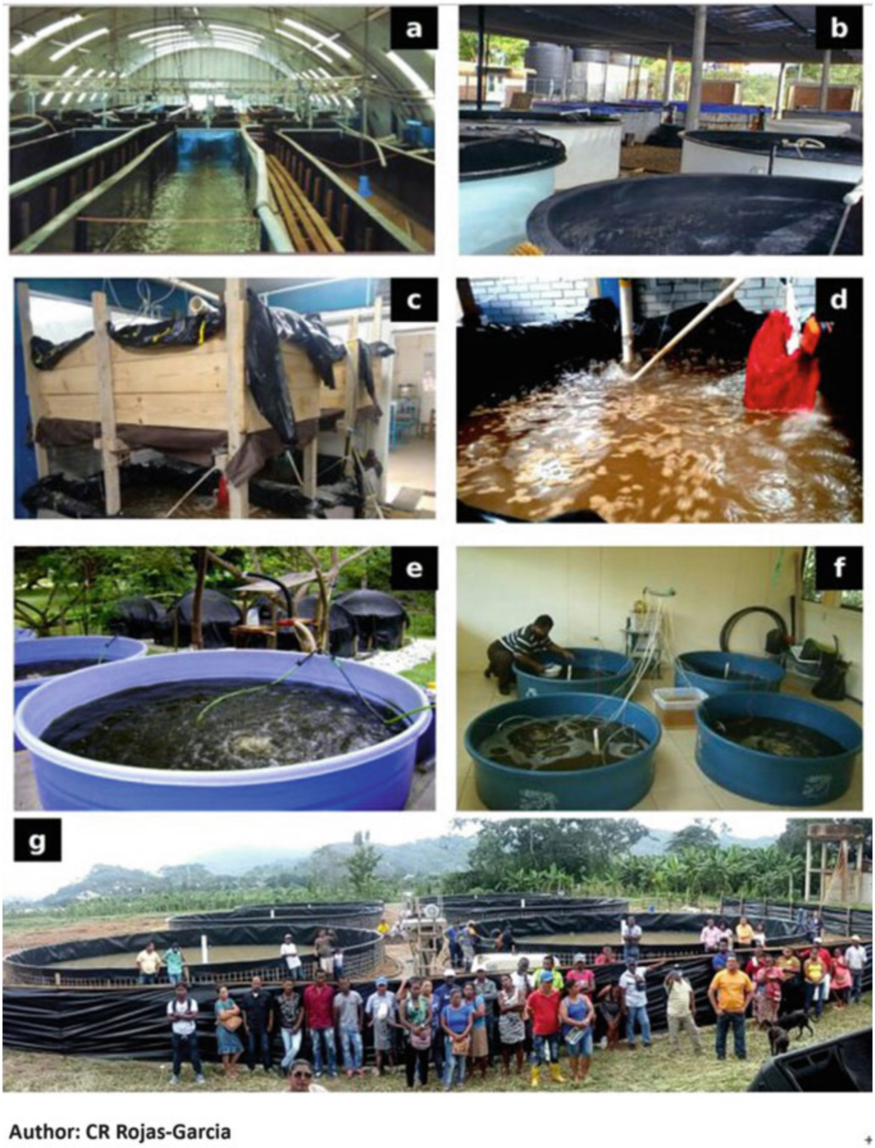


Fig. 7 Three decades of closed recirculation systems for shrimp farming in Mexico and Ecuador (1986–2017): **a** Sonora, Mexico, 1986; **b** Mazatlan, Mexico, 1996; **c**–**d** Veracruz, Mexico, 2012; **e** Manabi, Ecuador 2013; **f** Babahoyo, Ecuador, 2015 and; **g** Esmeraldas, Ecuador, 2018

1.9.5 Other Non-Viral Pathogenic Threats

Vibrio species: Genus *Vibrio* comprise the most frequent bacteria encountered in water environments. Bacterial pathogens of shrimp as *V. harveyi* [75, 84] possess virulence factors including various enzymes (e.g., proteases and lipases), siderophores, and proteinaceous toxins have been identified. White feces syndrome (WFS) is an emerging problem for penaeid shrimp farming industries in SE Asia countries, Thailand, Malaysia, Vietnam, Indonesia, China, and India. This occurrence of this syndrome is usually first evidenced by the appearance of white fecal strings floating on the surface of the shrimp ponds; within the white feces were found densely packed spores of the microsporidian Enterocytozoon hepatopenaei (abbreviated as EHP). Gills disease shrimp Impaired physiological functions in shrimp that are symptomatic of black gill. Melanized nodules, which can be associated with black gill, can form on the gills of shrimp and obstruct respiratory function and ion regulation.

1.9.6 Factors Affecting the Intestinal Microbiota - Developmental Stages

The ontogenic gut colonization in *L. vannamei* occurs at the nauplius stage when the anal pore begins a retrograde contractile movement for drinking water; this early colonization is previous to the mouth opening and exogenous feeding [84, 85]; retrograde peristalsis is an innate gut function also observed in larval fish [85, 86]. The intestinal bacterial community in *L. vannamei* varies with developmental stage and age. Huang [86, 87] used 454 pyrosequencing techniques to study the composition of the intestinal bacterial community in shrimp PL14 and one-, two- and three-month-old juvenile shrimp and found that three microbe phyla, namely Proteobacteria, Bacteroidetes, and Actinobacteria, dominate the gut community at all developmental stages. Proteobacteria are the core microorganism in *L. vannamei*. Firmicutes (e.g., genus *Bacillus*) are butyrate-producing microorganisms that use fiber as a carbon supply to provide short-chain fatty acids, which might be utilized by the host to promote the health of the intestinal mucosa [87, 88]. The abundance of Firmicutes will increase within the shrimp gut that are fed corn starch, suggesting that they are sensitive to the supply of carbohydrates within the diet [89].

1.9.7 Sustainability and Resilience Aspects of Green Shrimp Farming: Future and Ancient Tools

The urban gardens appeared in the USA during the oil and financial crises of the 1970s, while in Europe, a group of artists and activists begun the grassroots movement related to human well-being, community building, and improvement of everyday living spaces in communal houses [89, 90]. In the UK and the Netherlands thereafter surges an environmentalist counterculture [90, 91]. In recent years in Denmark, urban gardens were claimed to serve minorities [91, 92]. During the 1980s and 1990s, other regional social movements continued working in urban improvement, reclamation, and the creation of green areas. Nevertheless, urban farming is not an invention of new generations, regarding that before the modern hydroponics raising, ancient cultures developed similar agriculture systems. The Aztec people raised plants on "chinampas" floating rafts on the lakes bordering Tenochtitlan (today Mexico City). In China, floating ``frame fields' were used for growing water, spinach, and culinary herbs in southeast China [92, 93]. The paddy (dike-pond) rice system of China, dating back to the fifteenth century, evolved out of a need to control regular floods and rice cultivation [92, 94]. Concerning modern shrimp technology, it has been erroneously associated with expensive technology and equipment, only available for skilled, wealthy investors. That is not completely true for our shrimp microbiome tech (SMT) model, for which we speculate that more people will get involved in the years to come.

1.9.8 Farming Resilience and Food Sovereign in Green Cities

The demand for aquatic food is increasing globally, and despite the efforts to satisfy this growing demand, there is not a panacea for environmentally sustainable aquaculture [94, 95]. Instead of this, intensive aquaculture systems are growing, even claiming poverty alleviation of coastal communities in developing countries, but an expansion of traditional ponds over terrestrial crop areas predicts conflicts with other production systems [95, 96], while the remaining natural coastal areas are threatened by floods and other environmental hazards due to climate change.

The Green Revolution is now universally recognized (although still debated), accompanied by the Evergreen Revolution, referring to the idea that it is possible to use technology to increase productivity in perpetuity, without ecological harm [96–98]. Political ideas of food sovereignty cannot overcome the real resilience of local production. So far, food sovereignty has been in the hands of politicians leading to bad influences on the food industry due to corruption. The obscure fact of not being certain of what poisson is ingested when we eat industrial food has caused big concerns in citizens with ecological consciousness and caring to eat healthily. Modern technology, climate change threads, and future uncertainty are creating a work-frame for taking in our own hands the duty of food acquisition and food resilience, by learning how to cultivate our own food, and after transmitting and heredity, good knowledge to future generations can ensure the respect and protection of the planet. We want to empower people to acquire a value seafood resource that has been denied to the majority of citizens, "a fresh shrimp available all year-round." The changes in farming practices and people's good attitudes are unstoppable and must concern industrial shrimp: The complaints have built up a global demand to eradicate unhealthy practices, negligence, and omission for damages reparation; the solely official environmental normative is void without collectives and community involvement.

To date, the pioneering ventures of shrimp farming in temperate rich countries have been highly successful with respect to marketing and high profit (up to 20 times the normal value of Mexican shrimp) but have made little to create local value chains and social contributions. A full value chain for urban shrimp farming must include urban postlarval supplies, low fish protein/oil feeds, clean energy supply, local market commercialization, transportation, and delivery alternatives to the use of fossil fuel combustion vehicles, avoiding long-term refrigeration of harvested shrimp.

1.9.9 To Heredity Restored Mangrove Forest Instead of Abandoned Shrimp Pond

Educated shrimp farmers would participate with their own tools, as deep information on sustainable shrimp farming is feasible anyplace with no externalities. Including socio-economic integration, gender equity is solely achievable in a green city context. The urban shrimp farmers will not be fishermen, but normal citizens, artists, doctors, architects, preachers, and all sorts of people. Shrimp farming in green cities will be the heredity to future generations instead of abandoned shrimp ponds and deforested mangrove. Today, the abandoned shrimp ponds become into arid saltmarsh/mangrove land, valueless, that once abandoned are hardly restored. Regardless of technological breakthroughs, with mangrove land undervalued, cheap concessions, and governments sometimes ignoring good practices, it is necessary to eradicate some predatory practices of many shrimp farmers that find it easier to cut and run rather than to stay and manage. Modern aquaculture technologies make it unnecessary to build new ponds substituting natural wetlands areas, mainly saltmarsh and mangrove areas [98, 99], as recent closed-circulation methods allow farms to be built far away from the ocean and the coastline. Restored mangrove ecosystems can be deliberately designed and engineered to provide valuable ecosystem services, be adaptable to climatic changes, and develop platforms for educating non-specialists about both the successes and failures of restored mangrove ecosystems [99, 100].

1.9.10 Sustainable Shrimp Farming in Green Cities—Still Must Be Paid Back for Ecosystem Services?

The lack of business scruples and the absence of ecological sense seem to prevail in owners of gigantic “camaroneras” from Ecuador. They established an industry based on the atrocious premise of destroying and escape rather than staying and restoring the mangrove ecosystem. They live abroad as uprooted from their place of origin and the natural resources that made them very rich. Given their lack of roots in the ecosystem, they can no longer think or grasp they have a big debt for reparation in accordance with the huge ecological damage shrimp farming that has caused, not having any intention for their own free will.

The global mangrove extent is around 14 million ha, which provides a broad array of ecosystem services [100–103]. Also, mangroves are ecosystems for shrimp and fish nursery, migrating birds, and marine mammals [103–105]. Humankind benefits from a multitude of resources and processes provided by ecosystems, collectively known as ecosystem services (ES), a topic that has experienced a growing interest in the last two decades, as biologists and economists have tried to quantify these benefits to create management interventions [105–107]. Most of those services have negatively changed, and it would be worst in coastal and marine environments, where the loss of ES can be exacerbated by climate change [107, 108]. To face the diversity of resource management, tools such as integrated coastal zone management, ecosystem-based management, community-based coastal management, marine protected areas, marine spatial planning, and ocean zoning are now including or even focusing on the ES framework [108–110]. Considering this, as urban shrimp farming grows, it is possible to plan strategies to recover saltmarsh and mangrove areas where damages occurred, either by governance regulation measures or by urban farmers' own activism. With global trends for sustainable shrimp farming escalating, the traditional shrimp farming in large earthen ponds must reduce their size for production, allowing the re-conversion of ponds to natural wetlands supplying SE again, separately from the shrimp ponds or integrated to them to serve as natural biofilters. The natural shrimp fisheries should be protected and preserved for the new generations; thus, being necessary to develop instruments to integrate new generation shrimp farming to current policies, the civil society of future green cities should be included. However, shrimp farming laws for each country differ markedly, and past attempts to control international shrimp farming by “FAO Code of Conduct for Responsible Fisheries” [110, 111], from 1998; the “International Principles for Shrimp Farming” from 2006, and the WWF’s “Standards for Responsible Shrimp Farming” from 2011, did not achieve good results. Additionally, international regulations do not apply in many world regions where bad practices in shrimp aquaculture are common, operating with a high degree of impunity. Moreover, when local legislations denounce the negative impacts of irresponsible shrimp farming, there is often a lack of political will to enforce them. The new generation of educated shrimp farmers will enter to compete in the shrimp trade; they will hopefully use responsible practices, including restoration or creation of blue carbon ecosystem instead of being forced to costly practices due to negligence of old harmful practices. Empowering collectives (citizens, families, race equality, gender equality, old people, poor or rich families) and individual women and men (engineer, ecologist, teacher, poet, artist, policymakers, chefs, etc.) involved in the creation of green cities to disrupt bad farming practices and to promote environmental restoration.

1.9.11 Services Delivered by Mangroves and Other Wetlands—Need of Mapping?

Due to its aquaculture importance, shrimp became a sort of top predator on the mangrove chain system and coastal biodiversity, because of that, other fauna inhabiting mangrove has been neglected, including crabs, bivalve mollusks, and fish, locally important as ancestral food and for the local economy at Ecuador, Mexico, and other poor farming countries. This type of natural food intimate is linked to a healthy ecosystem, and its responsible use should be included in the new generation approaches.

The urban shrimp farmer of green cities must realize that wetlands provide a range of environmental services and processes, vital to the health of the planet [111, 112], including the natural nurseries to protect the genetic pool of shrimp. Some estimates on mangrove worldwide distribution indicate that more than 3 million ha have disappeared within only 25 years (1980–2005), with a degradation rate of 1% per year [112, 113]. The conversion of mangrove cover to develop aquaculture/agriculture industry has been documented as one of the main mangrove degradation drivers, particularly in Southeast Asia [113–115]. Thus, poverty and lack of ecological education in tropical coastal areas of developing countries are obstacles for the implementation of restoration initiatives, even when local communities often had better understanding of their surrounding environment compared to foreigner experts.

Although a wide range of participatory mapping approaches for ecosystem services assessments has been applied in different countries, the local participation could be reinforced with economic incentives, where the activism of local farmers could play an essential role. The mangrove–fishery linkage remains controversial, but in some cases, species such as Penaeid shrimp display a positive relationship between mangrove area and adjacent fisheries production (coastal and offshore), even when some analyzed facts point out that the extent of intertidal areas explained variability in prawn production better than mangrove area [115–117].

1.9.12 Policies and Economic Solutions for Future Payments of Ecosystem Services

Payments for Ecosystem Services (PES) is a term used to describe a range of resource management tools that create positive economic incentives for the conservation of ecosystems, with the aim of securing ecosystem services such as provision of fresh water, carbon storage, maintenance of biodiversity, and recreation. PES participants suggest that a lack of supportive policy frameworks in many countries is a major barrier [117, 118], and scaling up positive results of existing PES schemes may require a specific policy and legal framework. It is significantly vital, considering that transparency in funding, rational budgets, and mechanisms to substantiate the pursued goals, a very challenging purpose. Despite this, over four hundred PES schemes are documented for watershed services programs worldwide, with examples on each continent and in twenty-nine countries. These schemes have jointly supported the protection or rehabilitation of 365 million hectares of ecosystems' essential services, significantly forest ecosystems. PES schemes have emerged within the context of diverse legal frameworks, with different participation from governmental agencies through PES-specific legislation and policy strategies [118–120].

1.9.13 Frameworks, Policies and Community Actions

In 2017, following the registration of over 1400 voluntary commitments (VCs) associated with SDG14 at the UN Ocean Conference, it was identified that these VCs together address nine thematic areas. For purposes of follow-up and implementation of these commitments leading up to the 2020 UN Ocean Conference, nine corresponding communities were created. Among them the Community of Ocean Action for Mangroves arose defining several focal points, including generating multi-stakeholder interest and engagement in the Mangrove Community. The importance of mangroves to ocean health is mirrored within the 116 voluntary commitments submitted up to now, considering the restoration, rehabilitation, protection, and management of mangroves and associated ecosystems. These voluntary commitments of the Mangrove Community for Ocean Action reflect the vital role that mangroves play in supporting livelihoods of coastal communities, sustainable fisheries, resilience to the impacts of climate change, and sequestering atmospheric CO_2 [120, 26]. In addition to the diversity of international plans, mangrove voluntary commitments are set in forty-one countries; there, green cities' collectives must play a leading role.

2 Conclusions

Urban shrimp farming in green cities is feasible and available for anybody. This new generation of shrimp farming system suits the ecological requirements for any green city in the world. It does save water and have zero effluent pollution. The sustainability and resilience of green cities depend on their food system but must act as a microbial consortium that does cooperatively instead of competing with others. Ecosystem services provided by mangroves reach the inland towns. In this way keeping safe the superior genetic pool of natural shrimp in the wild, which is the base of shrimp aquaculture and, a requirement to protect the mangrove ecosystems. The urban shrimp microbiome tech (SMT) permits a year-round production of shrimp with the lowest energetic cost. The consumer's preferences for natural organic shrimp are guaranteed with urban farming; the rest depends on the national idiosyncrasies, attitudes, and willingness to pay for his consuming choices. The small shrimp size (5–8 g) is as tasty and nourishing as luxurious sizes (over 12 g).

3 Concluding Remarks

- Marine shrimp farming could be adapted to urban environments at any green city location; high technology standards are not an important issue.
- The shrimp microbiome tech (SMT) for urban farming fulfills and surplus any sustainability requirement: It does save water, eradicate effluent pollution, and protect aquatic biodiversity; it is inclusive socially—race and gender inclusive.
- The acquisition of a sustainable urban shrimp farming system will empower people of green cities to question industrial shrimp farming practices.
- Industrial shrimp farming cannot warrant the healthy quality of shrimp products. No more frozen shrimp added chemical preservation.
- Small shrimp sizes taste as good as big sizes and are highly nourishing too, but most importantly, it takes four times fewer resources to make them.
- The local food systems will benefit from fresh shrimp year-round with top quality antibiotic-free, pollution-free, pesticides-free, hormones-free.
- Green cities have the potential for the collective defense of food rights and free will to buy and consume locally produced food. Governmental regulations or food industry standards regulation does not apply.

- Local trust and collective attitude are enough to gain a local social license for urban shrimp farming.
- The capacitive for innovating and re-invent of the collective's inhabitants of green cities will make the difference for true food sovereignty against governmental planned or global corporative plans.

Conflict of Interest Statement The authors declare no interest conflicts. The Shrimp Microbiome Tech (SMT) model was created by CRG copyright in process.

References

1. Swisher ME (2016) An overview of small farm direct marketing. FCS7211. Gainesville, University of Florida Institute of Food and Agricultural Sciences. edis.ifas.ufl.edu/fy597
2. Scarborough P, Appleby PN, Mizdrak A et al (2014) Dietary greenhouse gas emissions of meat-eaters, fish-eaters, vegetarians and vegans in the UK, Vol 125, pp 179–19
3. Reynolds C (2003) Pelagic community assembly and the habitat template. Bocconea 16:323–339
4. FAO (2017) http://www.fao.org/fishery/statistics/en
5. Attramadal KJ, Truong TMH, Bakke I et al (2014) RAS and microbial maturation as tools for K-selection of microbial communities improve survival in cod larvae. Aquaculture 432:483–490. https://doi.org/10.1016/j.aquaculture.2014.05.052
6. Li E, Xu C, Wang X et al (2018) Gut Microbiota and its Modulation for Healthy Farming of Pacific White Shrimp *Litopenaeus vannamei*. Rev Fish Sci Aquac. https://doi.org/10.1080/23308249.2018.1440530
7. Hostins B, Wasielesky W, Decamp O (2019) Managing input C/N ratio to reduce the risk of acute Hepatopancreatic necrosis disease (AHPND) outbreaks in biofloc systems—a laboratory study. Aquaculture 508:60–65
8. Vadstein O, Mo TA, Bergh Ø (2003) Microbial interactions, prophylaxis and diseases, culture of cold-water marine fish. Blackwell Publishing Ltd, Oxford, pp 28–72
9. Vestrum RI, Attramadal KJK, Winge P et al (2018) Rearing water treatment induces microbial selection influencing the microbiota and pathogen associated transcripts of Cod (*Gadus morhua*) Larvae. Front Microbiol 9:851. https://doi.org/10.3389/fmicb.2018.00851
10. Vadstein O, Øie G, Olsen Y et al (1993) A strategy to obtain microbial control during larval development of marine fish. In: H Reinertsen, LA Dahle, L Jørgensen, K Tvinnerheim (eds) Proceedings of the 1st international conference on fish farming technology. Trondheim, Norway. 9–12 Aug 1993, Balkema, Rotterdam, pp 69–75
11. De Schryver P, Vadstein O (2014) Ecological theory as a foundation to control pathogenic invasion in aquaculture. ISME J 8:2360–2368. https://doi.org/10.1038/ismej.2014.84
12. Martins CIM, Eding EH, Verdegem M et al (2010) New developments in recirculating aquaculture systems in Europe: a perspective on environmental sustainability. Aquac Eng 43:83–93. https://doi.org/10.1016/j.aquaeng.2010.09.002
13. Badiola M, Mendiola D, Bostock J (2012) Recirculating aquaculture systems (RAS) analysis: main issues on management and future challenges. Aquac Eng 51:26–35. https://doi.org/10.1016/j.aquaeng.2012.07.004
14. Blancheton J, Attramadal K, Michaud L et al (2013) Insight into bacterial population in aquaculture systems and its implication. Aquac Eng 53:30–39. https://doi.org/10.1016/j.aquaeng.2012.11.009

15. Attramadal KJK, Øie G, Størset TR et al (2012) The effects of moderate ozonation or high intensity UV-irradiation on the microbial environment in RAS for marine larvae. Aquaculture 330–333:121–129
16. Burford MA, Thompson PJ, McIntosh RP et al (2004) The contribution of flocculated material to shrimp (*Litopenaeus vannamei*) nutrition in a high-intensity, zero-exchange system. Aquaculture 232:525–537
17. FAO (2011) Aquaculture topics and activities. Trends in aquaculture production. In: A Lowther, S Vannuccini (eds) FAO fisheries and aquaculture department
18. Hill C, Guarner F, Reid G et al (2014) The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol 11:506–514
19. Kunapongkiti P, Rongsayamanont C, Nayramitsattha P, Limpiyakorn T (2020) Application of cell immobilization technology to promote nitritation: a review. Environ Eng Res 25: 807–818
20. Blackburne R, Yuan Z, Keller J (2008) Partial nitrification to nitrite using low dissolved oxygen concentration as the main selection factor. Biodegradation 19:303–312
21. Daims H, Nielsen JL, Nielsen PH et al (2001) In situ characterization of Nitrospira-like nitrite-oxidizing bacteria active in wastewater treatment plants. Appl Environ Microbiol 67:5273–5284
22. Lücker S, Schwarz J, Gruber-Dorninger C et al (2015) *Nitrotoga*-like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants. ISME J 9: 708–720
23. Nowka B, Daims H, Spieck E (2015) Comparison of oxidation kinetics of nitrite-oxidizing bacteria: nitrite availability as a key factor in niche differentiation. Appl Environ Microbiol 81:745–753
24. Daims H, Lebedeva EV, Pjevac P et al (2015) Complete nitrification by *Nitrospira* bacteria. Nature 528:504–509
25. Van Kessel MAHJ, Speth DR, Albertsen M et al (2015) Complete nitrification by a single microorganism. Nature 528:555–559
26. Offre P, Spang A, Schleper C (2013) Archaea in biogeochemical cycles. Annu Rev Microbiol 67:437–457. https://doi.org/10.1146/annurev-micro-092412-155614
27. Winkler MK, Kleerebezem R, de Bruin LM et al (2013) Microbial diversity differences within aerobic granular sludge and activated sludge flocs. Appl Microbiol Biotechnol 97:7447–7458. https://doi.org/10.1007/s00253-012-4472-7
28. Ebeling JM, Timmons MB, Bisogni JJ (2006) Engineering analysis of the stoichiometry of photoautotrophic, autotrophic, and heterotrophic control of ammonia-nitrogen in aquaculture production systems. Aquaculture 257:346–358
29. VanWyk P (2006) Production of *L. vannamei* in recirculating aquaculture systems: management and design considerations. In: Proceedings of the 6th international conference on recirculation aquaculture, pp 38–47
30. Sunil S, Adav D, Jong L, Lai JY (2010) Enhanced biological denitrification of high concentration of nitrite with supplementary carbon source. Appl Microbiol Biotechnol 85:773–778
31. Dangcong P, Bernet N, Delgenes J-P, Moletta R (1999) Aerobic granular sludge—a case report. Water Res 33:890–893
32. Beun JJ, Heijnen JJ, Van Loosdrecht MCM (2001) N-removal in a granular sludge sequencing batch reactor. Biotechnol Bioeng 75:82–92
33. deKreuk MK, vanLoosdrecht MCM (2005) Simultaneous COD, nitrogen, and phosphate removal by aerobic granular sludge. Biotechnol Bioeng 90:761–769
34. Weber SD, Ludwig W, Schleifer KH, Fried J (2007) Microbial composition and structure of aerobic granular sewage biofilms. Appl Environ Microbiol 73:6233–6240

35. Verawaty M, Pijuan M, Yuan Z, Bond PL (2012) Determining the mechanisms for aerobic granulation from mixed seed of floccular and crushed granules in activated sludge wastewater treatment. Water Res 46:761–771
36. Zhou D, Niu S, Xiong Y et al (2014) Microbial selection pressure is not a prerequisite for granulation: dynamic granulation and microbial community study in a complete mixing bioreactor. Bioresour Technol 161:102–108
37. Liu Y, Tay JH (2004) State of the art of biogranulation technology for wastewater treatment. Biotechnol Adv 22:533–563
38. Dolinšek J, Lagkouvardos I, Wanek W et al (2013) Interactions of nitrifying bacteria and heterotrophs: identification of a Micavibrio-like putative predator of *Nitrospira* spp. Appl Environ Microbiol 79:2027–2037
39. Winkler MK, Kleerebezem R, de Bruin LM et al (2013) Microbial diversity differences within aerobic granular sludge and activated sludge flocs. App Microbiol Biotechnol 97:7447–7458
40. Yang W, Zheng Z, Lu K, et al (2019). Manipulating the phytoplankton community has the potential to create a stable bacterioplankton community in a shrimp rearing environment. Aquaculture 520:734789 https://doi.org/10.1016/j.aquaculture.2019.734789
41. Winkler MKH, Kleerebezem R, Khunjar WO et al (2012) Evaluating the solid retention time of bacteria in flocculent and granular sludge. Water Res 46:4973–4980
42. Attramadal KJK, Minniti G, Oie G et al (2016) Microbial maturation of intake water at different carrying capacities affects microbial control in rearing tanks for marine fish larvae. Aquaculture 457:68–72
43. Clavel J, Julliard R, Devictor V (2011) Worldwide decline of specialist species: toward a global functional homogenization? Front Ecol Environ 9(222–22):8
44. Ringø E, Van Doan H, Lee S et al (2020) Probiotics, lactic acid bacteria and bacilli: interesting supplementation for aquaculture. J Appl Microbiol 129:116–136. https://doi.org/10.1111/jam.14628
45. Olofsson M, Lamela T, Nilsson E et al (2012) Seasonal variation of lipids and fatty acids of the microalgae *Nannochloropsis oculata* grown in outdoor large-scale photo-bioreactors. Energies 5:1577–1592
46. Nath S, Matozzo V, Bhandari D, Faggio C (2019) Growth and liver histology of *Channa punctatus* exposed to a common biofertilizer. Nat Prod Res 33:1591–1598
47. Wang YC, Hu SY, Chiu CS, Liu CH (2019) Multiple-strain probiotics appear to be more effective in improving the growth performance and health status of white shrimp, *Litopenaeus vannamei*, than single probiotic strains. Fish Shellfish Immunol 84:1050–1058
48. Zuo Z-H, Shang B-J, Shao Y-C et al (2019) Screening of intestinal probiotics and the effects of feeding probiotics on the growth, immune, digestive enzyme activity and intestinal flora of Litopenaeus vannamei. Fish Shellfish Immunol 86:160–168
49. Du Y, Wang B, Jiang K et al (2019) Exploring the influence of the surface proteins on probiotic effects performed by *Lactobacillus pentosus* HC-2 using transcriptome analysis in *Litopenaeus vannamei* midgut. Fish Shellfish Immunol 87:853–870
50. Chen B, Peng M, Tong W, et al (2020) The quorum quenching bacterium *Bacillus licheniformis* T-1 protects zebrafish against *Aeromonas hydrophila* infection. Prob Antimicrob Prot 12:160–171. doi.org/https://doi.org/10.1007/s12602-018-9495-7
51. Kokou F, Makridis P, Kentouri M et al (2012) Antibacterial activity in microalgae cultures. Aquac Res 43:1520–1527
52. Rojas-Garcia CR (1992) The characterization of a seawater bio-filtration system for production of post-larval shrimps *Penaeus* sp [*Caracterización del sistema de biofiltración de agua marina para producción de post-larvas de camarón Litopenaeus sp en el CINVESTAV-Mérida, Yucatán, México*]. BSc. Thesis UNAM, Class: 001-50322-R1-1992-2, 50 pp

53. Rojas-Garcia CR (1992) The effect of excluding vitamins and minerals premix from shrimp diets [*Efecto de exclusión de vitaminas y minerales en dietas experimentales para camarón (Penaeus vannamei) en cultivo semi-intensivo*] MSc Thesis, Cinvestav-Merida Call num MERI R65, 91 pp
54. Rojas-García CR (2008) Assessing quality attributes of plant proteins for fish feeding: protein digestion with different enzymes models. First Inter CGNA workshop, Temuco Chile, 4–6 Aug
55. Rojas-García CR (2009) Full plant protein based aqua-feeds: evaluation in juvenile aquatic species. Annual meeting of the World Aquaculture Society, Veracruz Mexico, 25–29 Sept
56. Rojas-García CR, Cabezas M, Mosquera M (2015) Experimental rearing of *Litopenaeus vannamei* postlarvae in shallow water and low salinity: adding high doses of beneficial bacteria makes any difference? Meeting FENACAM & LACQUA/SARA (WAS) '15, Fortaleza, Brazil, 16–19 Nov
57. Royal Gazette (2007) Vol 124 Section 78E, dated 29 June B.E. 2550. Available at:https://www.acfs.go.th/standard/download/eng/Organic_Marine_Shrimp.pdf
58. FAO (2019) http://www.fao.org/fishery/statistics/en (2019)
59. FAO (2007) Part 1 Biosecurity principles and components. In FAO Biosecurity Toolkit. Available from: http://www.fao.org/3/a1140e/a1140e01.pdf
60. Krogdahl A, Penn M, Thorsen J et al (2010) Important antinutrients in plant feedstuffs for aquaculture: an update on recent findings regarding responses in salmonids. Aquac Res 41:333–344
61. Krogdahl A, Gajardo K, Kortner TM et al (2015) Soya saponins induce enteritis in Atlantic salmon (*Salmo salar* L). J Agric Food Chem 63: 3887–3902
62. Grammes F, Reveco FE, Romarheim OH, et al (2013) *Candida utilis* and *Chlorella vulgaris* counteract intestinal inflammation in Atlantic salmon (*Salmo salar* L.). PLoS One 8:e83213. doi:https://doi.org/10.1371/journal.pone.0083213
63. Espe M, Lemme A, Petri A et al (2006) Can Atlantic salmon (*Salmo salar*) grow on diets devoid of fish meal? Aquaculture 255:255–262. https://doi.org/10.1016/j.aquaculture.2005.12.030
64. Samtiya M, Aluko RE, Dhewa T (2020) Plant food anti-nutritional factors and their reduction strategies: an overview. Food Prod Process Nutr 2:6. https://doi.org/10.1186/s43014-020-0020-5
65. Hua K et al (2019) The future of aquatic protein: implications for protein sources in aquaculture diets. One Earth 1:316–329
66. Francis G, Makkar HPS, Becker K (2001) Effects of *Quillaja saponins* on growth, metabolism, egg production and muscle cholesterol in individually reared Nile tilapia (*Oreochromis niloticus*). Comp Biochem Physiol C Toxicol Pharmacol 129:105–114
67. Budhwar S, Sethi K, Chakraborty M (2020) Efficacy of germination and probiotic fermentation on underutilized cereal and millet grains. Food Prod Process Nutr 2:12. https://doi.org/10.1186/s43014-020-00026-w
68. Palmegiano GB, Agradi E, Forneris G et al (2005) Spirulina as a nutrient source in diets for growing sturgeon (*Acipenser baeri*). Aquac Res 36:188–195
69. Becker EW (2007) Micro-algae as a source of protein. Biotechnol Adv 25:207–210
70. Kotrbacek V, Doubek J, Doucha J (2015) The chlorococcalean alga *Chlorella* in animal nutrition: a review. J Appl Phycol 27:2173–2180
71. Tonheim SK, Espe M, Hamre K, Rønnestad I (2005) Pre-hydrolysis improves utilisation of dietary protein in the larval teleost Atlantic halibut (*Hippoglossus hippoglossus* L.). J Exp Mar Biol Ecol 321:19–34
72. Drossou A, Ueberschär B, Rosenthal H, Herzig K (2006) Ontogenic development of the proteolytic digestion activities in larvae *Oreochromis niloticus* fed with different diets. Aquaculture 256:479–488
73. Mersch J, Beauvais MN, Nagel P (1996) Induction of micronuclei in haemocytes and gill cells of zebra mussels, *Dreissena polymorpha*, exposed to clastogens. Mutat Res 371:47–55

74. Gatesoupe FJ (2010) Probiotics and other microbial manipulations in fish feeds: prospective health benefits. In: Watson RR, Preedy VR (eds) Bioactive Foods in Promoting Health. Academic Press, Oxford, pp 541–552
75. Romanova E, Spirina E, Romanov V, et al (2020). Effects of *Bacillus subtilis* and *Bacillus licheniformis* on catfish in industrial aquaculture. E3S Web of Conf 175:02013. doi:https://doi.org/10.1051/e3sconf/202017502013
76. Vince J, Haward M (2019) Hybrid governance in aquaculture: Certification schemes and third party accreditation. Aquaculture 507:322–328. https://doi.org/10.1016/j.aquaculture.2019.04.041
77. Vormedal I, Gulbrandsen LH (2020) Business interests in salmon aquaculture certification: competition or collective action? Regul Governance 14:328–343. https://doi.org/10.1111/rego.12213
78. Vormedal I, Skjærseth J (2020) The good, the bad, or the ugly? Corporate strategies, size, and environmental regulation in the fish-farming industry. Bus Polit 22:510–538. https://doi.org/10.1017/bap.2019.30
79. Zorriehzahra MJ, Delshad ST, Adel M et al (2016) Probiotics as beneficial microbes in aquaculture: an update on their multiple modes of action: a review. Vet Q 36:228–241. https://doi.org/10.1080/01652176.2016.1172132
80. Soltani M, Lymbery A, Ghosh K et al (2019) Genus *Bacillus*, promising probiotics in aquaculture: Aquatic animal origin, bio-active components, bioremediation and efficacy in fish and shellfish. Rev Fish Sci Aquac 7(3):331–379. https://doi.org/10.1080/23308249.2019.1597010
81. Joffre OM, Klerkx L, Khoa TND (2018) Aquaculture innovation system analysis of transition to sustainable intensification in shrimp farming. Agron Sustain Dev 38:34. doi.org/https://doi.org/10.1007/s13593-018-0511-9
82. Thitamadee S, Prachumwat A, Srisala J et al (2016) Review of current disease threats for cultivated penaeid shrimp in Asia. Aquaculture 452:69–87
83. Turkmen G, Toksen E (2010). Biosecurity and major diseases in shrimp culture. In: 2nd International symposium on sustainable development, June 8–9. Sarajevo
84. Li E, Xu C, Wang X et al (2018) Gut microbiota and its modulation for healthy farming of Pacific White Shrimp *Litopenaeus vannamei*. Rev Fish Sci Aquac 26:381–399. https://doi.org/10.1080/23308249.2018.1440530
85. Rønnestad I, Rojas-García CR, Skadal J (2000) Retrograde peristalsis, a possible mechanism for filling the pyloric caeca? J Fish Biol 56:216–218
86. Huang Z, Li X, Wang L, Shao Z (2016) Changes in the intestinal bacterial community during the growth of white shrimp. *Litopenaeus vannamei.* Aquac Res 47:1737–1746
87. Velazquez OC, Lederer HM, Rombeau JL (1997) Butyrate and the colonocyte: production, absorption, metabolism, and therapeutic implications. Oxygen Transport Tissue XXXIII 427:123–134
88. Qiao F, Liu YK, Sun YH et al (2017) Influence of different dietary carbohydrate sources on the growth and intestinal microbiota of *Litopenaeus vannamei* at low salinity. Aquacult Nut 23:444–452
89. Ioannou B, Morán N, Sondermann M et al (2015). Grassroots gardening movements: towards cooperative forms of green Urban development? In: Urban allotment gardens in Europe, Routledge
90. McKay G (2011) Radical gardening: politics, idealism and rebellion in the garden. Frances Lincoln, London
91. Jarvis I, Gergel S, Koehoorn M, vandenBosch M (2019) Greenspace access does not correspond to nature exposure: measures of urban natural space with implications for health research. Landscape and Urban Planning 194. https://doi.org/10.1016/j.landurbplan.2019.103686
92. Simoons FJ (1990) Food in China: a cultural and historical inquiry. CRC Press, Spokane, WA USA, p 140

93. Adams REW (2005) Prehistoric mesoamerica. University of Oklahoma press, Norman, OK USA, p 544
94. Klinger D, Naylor R (2012) Searching for solutions in aquaculture: charting a sustainable course. Annu Rev Environ Resour 37:247–276
95. Rakocy JE, Masser MP, Losordo TM (2006) Recirculating aquaculture tank production systems: aquaponics—integrating fish and plant culture. South Reg Aquacult Cent 54:1–16
96. Swaminathan MS (2004) Ever-green revolution and sustainable food security. In: A Eaglesham, A Wildeman RWF Hardy (eds) Agricultural biotechnology: finding common international goals. National Agricultural Biotechnology Council, Ithaca
97. Sumberg J, Thompson J (2013) Revolution reconsidered: evolving perspectives on livestock production and consumption, STEPS working paper 52. STEPS Centre, Brighton
98. Berlanga-Robles CA, Ruiz-Luna A, Hernández-Guzmán R (2011) Impact of shrimp farming on mangrove forest and other coastal wetlands: the case of Mexico. In: Sladonja B. (ed) Aquaculture and the environment—a shared destiny. Chapter 2:17–30. InTech Open Science
99. Ellison AM, Felson AJ, Friess DA (2020) Mangrove rehabilitation and restoration as experimental adaptive management. Front Mar Sci 7:327. https://doi.org/10.3389/fmars.2020.00327
100. Barbier EB, Hacker SD, Kennedy C et al (2011) The value of estuarine and coastal ecosystem services. Ecol Monogr 81:169–193. https://doi.org/10.1890/10-1510.1
101. Giri C, Ochieng E, Tieszen LL et al (2011) Status and distribution of mangrove forests of the world using earth observation satellite data. Glob Ecol Biogeogr 20:154–159
102. Sanderman J, et al (2018) A global map of mangrove forest soil carbon at 30 m spatial resolution. Environ Res Lett 13 055002. doi:https://doi.org/10.1088/1748- 9326/aabe1c
103. Mumby PJ, Edwards AJ, Arlas-Gonzalez JE et al (2004) Mangroves enhance the biomass of coral reef fish communities in the Caribbean. Nature 427:533–536
104. Nagelkerken I, Bothwell J, Nemeth RS, Pitt JM, vanderVelde G, (2008) Interlinkage between Caribbean coral reefs and seagrass beds through feeding migrations by grunts (Haemulidae) depends on habitat accessibility. Mar Ecol Progr 368:155–164
105. Sikor T (ed) (2013) The justices and injustices of ecosystems services edited. Routledge, New York USA
106. Costanza R (2019) Ecological economics 1. In: Fath B (ed) Encyclopedia of ecology. Elsevier, Amsterdam, pp 258–26474
107. Lau WWY (2012) Beyond carbon: conceptualizing payments for ecosystem services in blue forests on carbon and other marine and coastal ecosystem services. Ocean Coast Manag. https://doi.org/10.1016/j.ocecoaman.2012.03.011
108. Agardy T (2010) Ocean zoning: making marine management more effective. Earthscan, Washington
109. Granek EF, Polasky S, Kappel CV et al (2009) Ecosystem services as a common language for coastal ecosystem-based management. Conser Biol 24:207–216. https://doi.org/10.1016/j.marpol.2016.09.021
110. FAO (1995) Code of conduct for responsible fisheries. Rome, p 41
111. Stenton-Dozey J, Broekhuizen N (2019) Provisioning of ecological and ecosystem services by shrimp: a literature review in context of the state of the environment pre-and post-mussel farming. NIWA Client Report No: 2019020H. 141
112. Mayaux P, Holmgren P, Achard F et al (2005) Tropical forest cover change in the 1990s and options for future monitoring. Philos Trans Royal Soc London B, Biol Sci 360:373–384. https://doi.org/10.1098/rstb.2004.1590
113. Thomas N, Lucas R, Bunting P et al (2017) Distribution and drivers of global mangrove forest change, 1996–2010. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0179302
114. Damastuti E, de Groot R (2019) Participatory ecosystem service mapping to enhance community-based mangrove rehabilitation and management in Demak, Indonesia. Reg Environ Change 19:65–78. https://doi.org/10.1007/s10113-018-1378-7

115. Lee SY (2004) Relationship between mangrove abundance and tropical prawn production: a re-evaluation. Mar Biol 145:943–949
116. Carrasquilla-Henao M, Juanes F (2017) Mangroves enhance local fisheries catches: a global meta-analysis. Fish Fish 18:79–93
117. Bennett NJ, Dearden P, Murray G, Kadfak A (2014) The capacity to adapt? Communities in a changing climate, environment, and economy on the northern Andaman coast of Thailand. Ecol Soc 19:5. https://doi.org/10.5751/ES-06315-190205
118. Van Hecken G, Bastiaensen J (2010) Payments for ecosystem services: justified or not? A political view. Environ Sci Policy 13:785–792. https://doi.org/10.1016/j.envsci.2010.09.006
119. Jackson S (2015) Legal frameworks for payments for ecosystem services. Comparative policy approaches to establishing, regulating and enabling payments to conserve ecosystems, Vol 7 IUCNAEL EJournal
120. UN-DESA (2019) The community of ocean action for mangroves—towards the implementation of SDG14 interim report

Carlos R. Rojas-Garcia is a Marine Biologist, co-founder of CIAD-Mazatlan (1993–1997) and developer of recirculating aquaculture systems (RAS) for shrimp rearing and fish nutrition research. Since 1997, he became part of the science diaspora, achieving a recognized scientific career abroad, working for several research groups of Scandinavia, Europe, and South America. His research interests are the development of digestive function of larval aquatic organisms, RAS technology with microbiome tech and sustainable farming of crustaceans mainly shrimp and freshwater prawns. From 2013 to 2017, he worked for the Ministry of Education Science and Technology of Ecuador to execute the technology transfer for closed aquaculture of shrimp and native species changing the old production matrix of shrimp and developing the biosecure system for postlarval nursery that has eradicated the threat of devastating diseases.

Isabel Jimenez is a biologist, with a Master of Science (Animal Biology), from the National Autonomous University of Mexico, Mexico City, and has doctoral studies at the Center for Research and Advanced Studies -SEP at Merida. Since 2008, she has worked at the Tecnológico Nacional de México/Instituto Tecnológico de Boca del Río, Veracruz, Mexico, where she has worked as a Teacher and Researcher in the postgraduate course in Aquaculture (Laboratory of Aquatic Parasitology). Dr. JImenez conducted research on the temporal behavior of infectious diseases caused by helminths and protozoa of wild freshwater fish, as well as in tilapia hatcheries cultured in recirculation systems, and under laboratory conditions. Additionally, she has worked on morphological and molecular identification of parasites with pathogenic potential in tilapias and some commercially important invertebrates.

17 Public Understanding of Geoethics Dilemmas: A survey on Lithium Exploitation

Clara Vasconcelos, Sara Moutinho, Tiago Ribeiro, and Alexandre Lima

I have learned to be open to surprises (…) to not have preconceived ideas or close your mind from listening to what might work.

John B. Goodenough.

2019 Nobel prize winner

C. Vasconcelos (✉) · T. Ribeiro
Interdisciplinary Centre of Marine and Environmental Research of the University of Porto; Departamento de Geociências, Ambiente e Ordenamento Do Território/Unidade de Ensino das Ciências, Faculdade de Ciências, Universidade Do Porto, Rua Do Campo Alegre S/N, 4169-007 Porto, Portugal
e-mail: csvascon@fc.up.pt

T. Ribeiro
e-mail: tiago.ribeiro@fc.up.pt

C. Vasconcelos · T. Ribeiro
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Porto, Portugal

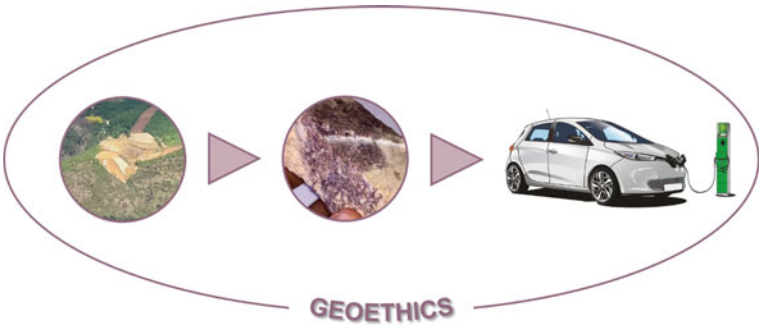
S. Moutinho
Instituto de Ciências da Terra, Porto, Portugal
e-mail: sara.moutimho@fc.up.pt

A. Lima
Departamento de Geociências, Ambiente e Ordenamento Do Território, Faculdade de Ciências, Instituto de Ciências da Terra, Universidade Do Porto, Rua Do Campo Alegre S/N, 4169-007 Porto, Portugal
e-mail: allima@fc.up.pt

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_17

Summary

Lithium holds high potential in technology, mainly green technologies, presenting notable economic and environmental importance. In the north of Portugal, there are several regions with particular potential for lithium exploration. Portuguese exploitation of lithium has generated significant controversies, mainly in the populations neighboring the potential exploitation areas. To understand Portuguese citizens' perceptions about the exploitation of metallic minerals in Portugal, particularly lithium, and their opinions on geoethical dilemmas associated with this activity, a questionnaire was applied to 132 participants, aged between 18 and 74 years (average of 33.6 years), of which 56 (42.4%) were female and 76 (57.6%) male. The majority of participants ($n = 118$; 89.39%) recognize the potential of lithium in the production of batteries for electronic devices and cars, as well as the locations of Covas do Barroso and Sepeda as two of the main sites for lithium exploitation in Portugal ($n = 49$; 37.12%). Most participants consider that a geoethical reflection should be made regarding the exploration and exploitation of lithium, the location of nuclear waste safe deposits, the sustainable infrastructures' construction, the development of geoparks, and organization of geological tours/hikes in areas of geological and landscape interest. They also admit that it is essential to adopt behaviors and attitudes related to ethical, cultural, and social values when carrying out prospecting and mining.



Geoethics

Lithium-ion batteries are growing in popularity and are often used in electric vehicles as green energy.

The code of this chapter is 01001100 01101001 01110100 01101000 01101001 01110101 01101101 00001101 00001010.

1 Introduction

The periodic table celebrated 150 years in 2019. This table's original version dates back to 1869, published by the chemist Dmitri Mendeleev (1834–1907). It contained the 63 chemical elements known. His work consisted of organizing the various chemical elements according to their periodic trends and properties: He arranged elements inline in ascending atomic mass order and by columns the elements with similar properties, having noticed that there was a network of vertical, horizontal, and diagonal relationships between them [1]. This instrument, which makes it possible to predict Earth's elements' properties, represents a turning point in science history. It consists of the most objective, explicit, and accurate tool for organizing all chemical elements known. One of the relevant chemical elements, lithium, has been the subject of controversy and geoethical dilemmas. Its exploration in Portugal is currently being discussed, both by the scientific community and by society, mainly related to Portugal's potential to exploit this element. Lithium is widely used today, for example, in batteries for electronic devices and electric cars. In Portugal, in recent years, some regions' potential for lithium exploitation has been explored, having obtained exciting results, much-discussed by the media, the municipalities, and, mainly, by the inhabitants. In this context, geoethics can make a valuable contribution by clarifying citizens and promoting ethical, cultural, and environmental reflection by lithium exploration and exploitation companies, reflecting on the values that must sustain behavior and scientific practice whenever human activities relate to the geosphere.

In the present study, we started from the research question, what is the public understanding of metallic minerals and their applications, particularly lithium? We also sought to determine citizens' opinions on the importance of geoethical reflection in exploiting metallic minerals. For this purpose, we carried out a survey to the local community where lithium prospecting took place with a view of its possible exploitation.

2 Geoethics: A New but Consolidated Study Area

Geoethics emerged in geosciences to reflect on those *values that underpin appropriate behaviors and practices wherever human activities interact with the Earth system*. It can be considered that geoethics arose when geoscientists began to be aware that their activities when interacting with the geosphere could irreversibly change the Earth's natural processes and that their choices could have would be harmful to the environment, affecting species and their survival, including the human species [2].

Geoethics is an emerging but already consolidated branch of knowledge in geosciences that relates scientific content to the technological, methodological, and sociocultural aspects of Earth sciences and ethics. This discipline represents a point of intersection between humanistic disciplines, such as philosophy, epistemology, sociology and ethics, and scientific fields, such as geology, natural sciences and ecology, and economics [3]. According to the International Association for Promoting Geoethics (IAPG), geoethics refers to the research and attempts to reflect on the values on which behaviors and practices should be based whenever human activities intercross the geosphere [4]. Geoethics can be considered a widely articulated discipline with several fields of knowledge, dealing, for example: with the conservation, the exploration of geological resources and their exploitation, the ecosystems protection, research and education in geosciences, geological risk management and mitigation, and the geodiversity promotion [3].

Given that their work is, most of the time, directly related to society, geoscientists, namely geologists, should reflect on the communication of their discoveries and uncertainties [5], trying to do so in a transparent, objective, and straightforward way [6] so that all citizens understand them. In this sense, geoethics also earn the communication ethics status for being concerned with the relationship/interaction between science and the public [6].

The geologists' professional activities and other geoscientists are frequently related to the territory, operating as a dynamic system. Therefore, they must be reflected and thought about in the long-term within the legislative framework. Matteucci and collaborators [7] give some examples:

- "Assessment of the sustainability and long-term effects of removing material from the riverbed;
- Evaluation of a location for storage of material or radioactive waste suitability;
- Stabilization of a landslide through expensive structural measures or "natural" drainage and probably low-cost systems;
- Land-use planning; and
- Geological heritage and landscape conservation."

Currently, there seems to be an awareness that a dynamic approach to achieving balanced territory management must be based on knowledge shared between geoscientists and citizens, involving society in a participatory democracy [8]. However, for citizens to participate actively in the management of the territory, contributing constructively to decision making, they must be well informed and aware of the technical and scientific issues related to the subject under discussion. Now, geoscientists have a central role at this level. As specialists in the scientific contents, they must transmit and explain scientific information to citizens in a language they can understand.

Geologists, and geoscientists in general, are increasingly committed to bringing their skills and know-how about the problems surrounding the environment to society. The participation of specialists in discussion groups (debates, for example) is increasingly frequent, and their availability to clarify doubts and advise citizens concerning scientific content.

Given that geoethics is a branch of knowledge based on a reflection on human beings' activities over the environment, to adopt appropriate behaviors and practices in this domain, a set of ethical, cultural, and social reference values must be defined on which stakeholders should base their decisions. These values should guide the choice or decision to be made concerning a specific situation [2, 5, 9]. There are multiple geoethical works published regarding mining, namely The *White Paper on Responsible Mining* of the IAPG [10]. According to Boon [11], there is a serious area of codes and guidelines including: *"transparent and ethical behavior; compliance with applicable law; consistency with international norms of behavior; taking into account the expectations of stakeholders; integration throughout the organization; and practicing these in the organization's relationships."* According to the same author, various alternative conceptions to aid this situation can be described as: *"(i) The Social and Environmental Value Governance Ecosystem (SEVGE) approach views corporations as only one of many actors in a complex ecosystem with environmental and social values at its core and overall system governance taking place through a variety of mechanisms, one of which is Creating Shared Resource(CSR) (…), (ii) Creating Shared Value (CSV) is defined as policies and operating practices that enhance the competitiveness of a company while simultaneously advancing the economic and social conditions in the communities in which it operates (…); (iii) Integral Social Responsibility that applies to all actors (…) governments and communities as well as companies have social responsibilities"* [11, p. 10]. Such recommendations are also worth considering when exploiting lithium in Portugal or anywhere else in the world.

3 Lithium Exploitation: A SURVEY

Given the significant volume of lithium reserves and their value in Portugal, a survey was carried out based on a questionnaire administered to 132 Portuguese citizens that combined questions about geoethical dilemmas of lithium exploration in regions of the North and Center of Portugal.

3.1 Lithium Characteristics and Properties

In Mendeleev's periodic table, lithium (Li) is the third element with an atomic mass of 6.941. This element consists of three protons and three electrons. Lithium belongs to the alkali metal family (group I of the periodic table). However, it is different from sodium, potassium, rubidium, and cesium in its ionic radius being smaller than the other elements of the same group and presenting a high electrochemical potential [12].

Lithium has a low density than other elements (0.53–20 °C), solid at room temperature. Given its density, lithium is a light element with the ability to float in water. While floating in the water, lithium can react with it, releasing hydrogen from this chemical reaction [12, 13]. As its ionic radius is lower than that of the other elements in group I of the periodic table, lithium cannot replace alkaline ions with larger dimensions. During the late phases of magmatic crystallization, lithium, however, can partially replace iron (Fe), aluminum (Al), and magnesium (Mg) in some minerals if particular conditions are met.

In a medium to high geological environment, lithium binds preferentially to silicates and not to sulfides or metals [12]. Silicates frequently occur in silicate minerals such as spodumene (lithium and aluminum silicate), montebrasite-amblygonite (lithium and aluminum phosphate), lepidolite (phyllosilicate), and petalite (tectosilicate) (Fig. 1).

Lithium can be used directly in mineral concentrate, metal, or various chemical by-products (carbonate, hydroxide) obtained from brines or lithium minerals [14]. In addition to its recognized chemical and geological value, lithium also plays a crucial role in regulating brain function and treating bipolar disorders and depressions [13]. The benefits of lithium on the human organism's balance and regulation have remained unknown for a long time. However, some curious data has been

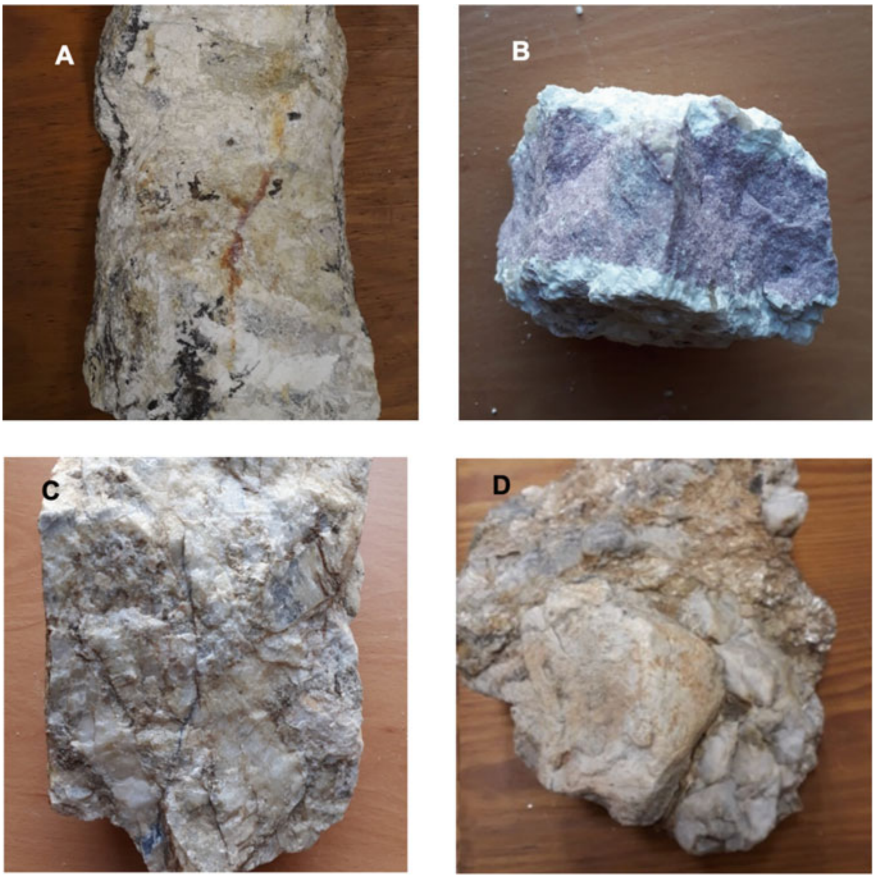


Fig. 1 Minerals of spodumene (**a**), lepidolite (**b**), petalite (**c**), and montebrasite (**d**) (photography by Alexandre Lima)

gathered. For example, lithium-treated patients might have increased white blood cell (granulocyte) count and decreased blood lymphocyte counts. Some vets administer lithium to animals after chemotherapy and radiation treatments that reduced leukocyte counts. Also, lithium helps restore white blood cells count more quickly [15].

The potential of lithium in the production of batteries for electric cars and electronic devices is widespread. However, the potential of lithium extends to other uses, more or less explored and publicized. Christmann and collaborators [12] presented a percentage distribution of lithium's main benefits, stating that its primary use concerns the batteries production. However, it is common to use them in the manufacture of ceramics and glass. It is also frequent in obtaining lubricating oils, and this application corresponded in 2013 to the third most significant lithium application and corresponding to 3885 tons.

3.2 Lithium Exploitation in Portugal

After describing lithium's main characteristics and properties and its main uses, how lithium exploration has developed in Portugal remains to be explained. In recent years, interest in lithium exploitation has increased slightly everywhere, but several prospecting studies have been designed to determine Portugal's potential for exploiting this element.

In their work, Carvalho and Farinha [15] indicated the main places in Portugal where lithium-rich minerals are found (Fig. 2). However, there are more locations in Portugal where lithium minerals can be found: Serra de Arga; Covas do Barroso; Barca d'Alva; Guarda; Mangualde; Segura. According to the same authors, lithium's properties tend to be concentrated in more differentiated granitic magmas, usually formed after its consolidation. This type of granitic rock occurs, more often, in the North and the Center of Portugal, namely in the areas of Galicia-Trás-os-Montes and in the Center-Iberian zone. In northern Portugal, one of the best-known areas is Covas do Barroso (Vila Real district), in the Barroso-Alvão region. This region is characterized by the presence of a large number of pegmatitic and aplite-pegmatitic veins on granite. In the Barroso region, in the village of Alijó, the pegmatites exploitation with lithium (Fig. 2) for the ceramics industry is currently underway by Lagoa Group. The Portuguese case has been discussed, compared to the panorama extended to other countries globally, given the lithium consumption increase, which has registered a significant rise in recent years.

Lithium production has been done mostly in countries like Chile, Australia, and Argentina [14]. Although lithium's main sources are litiniferous brines, spodumene is a lithium ore, namely in China, Australia, and Brazil. Moreover, it will probably return to be in the USA, Canada, and Russia [16]. Considering that Portugal's spodumene reserves are considered equivalent, for example, to those in Finland, where the first lithium mine is starting [16], some specialists consider that the Portuguese exploitation should also be studied and promoted in obtaining lithium for batteries. However, this is a subject widely discussed and reported by the media,

Fig. 2 Exploitation of pegmatites with lithium in the village of Alijó (Barroso region), by the Lagoa Group (photography by Alexandre Lima)

which conveys great controversy and discord on the part of the populations, mainly those born in the main places of exploitation. Thus, a deep reflection is justified, not only from a scientific point of view but also from society's contribution.

4 Methodology

We carried out a survey using a questionnaire built and determined its validity and reliability. The analysis of the data was based, mainly, on a descriptive statistical treatment, having been calculated, using the statistical software SPSS (version 25), the frequencies of the answers in each question, with the respective values of percentages, but also average values, for example, the ages of the participants. A chi-square independence test was also performed to ascertain significant relationships between some variables.

4.1 The Questionnaire

This study aimed to survey Portuguese citizens' perceptions about the use of metallic minerals in Portugal and the importance of a geoethical reflection associated with these activities. A questionnaire was developed with specific questions to analyze each of these aspects. Consequently, the constructed instrument was composed of nine questions with a direct answer, more specifically, multiple-choice questions and some on a Likert scale (Appendix 1). The first five questions (Q1–Q5) are related to the mineral concept and to the metallic minerals' identification and minerals that have lithium in their constitution. Besides, they also relate to the exploration of lithium in Portugal, namely, the main exploitation sites and their applications. For this purpose, all the questions asked were multiple-choice, with the participants had to select three options in each of them. To analyze the Portuguese citizens' opinions on the importance of geoethical reflection applied to the mining activity, four questions were asked (Q6–Q9), two multiple-choice (Q6 and Q7), and two Likert scales (Q8 and Q9). In question Q6, participants must select the option that most closely matches the definition of geoethics, possibly indicating their conception/definition. Question Q7 focuses on the geoethical values and, for that, ten statements were presented that the participants must relate to these, selecting which of the three values represented: ethical, cultural, and social values. In question Q8, it was intended that the participants indicate their degree of agreement, on a scale of 1–4: 1—disagree entirely; 2—disagree; 3—agree; and 4—agree entirely, regarding several situations that deserve geoethical reflection. Finally, in question Q9, the participants are confronted with ten statements that reflect behaviors/attitudes based on the main values of geoethics that must be adopted when mining and indicate the degree of importance they assign to each one. The degree of importance was measured on a scale of 1–6: 1—not important at all; 2—not very important; 3—important; 4—very important; 5—quite important; and 6—extremely important. In addition to these questions directly related to the objectives of the work carried out, sociodemographic data were also collected, which may assist in the analysis and interpretation and help in the study sample's characterization. These data include gender, age (until December 2019), academic qualifications, district of residence, and employment. Before its implementation, three specialists validated the questionnaire, whose areas of intervention are geosciences, specifically prospecting and mining, and geoethics. In a pilot study, the questionnaire was applied to some individuals from the population who were not part of the study sample to assess the instrument's fidelity.

4.2 Sample

This study sample consisted of 132 volunteer participants who answered the questionnaire through the Faculty of Sciences of Porto University webmail or social networks. Table 1 summarizes some sociodemographic data of the participants.

The analysis of Table 1 shows that the sample comprised 56 females (42.4%) and 76 males (57.6%), most of whom have higher education studies, which can be a degree, an MSc, or a Ph.D. (n = 107; 81.1%). Most respondents were from the North of the country. Regarding their professional situation, each participant indicated in the questionnaire what their profession/professional situation was at the moment, as shown in Table 1.

5 Results and Discussion

Given the diversity of results obtained with the questionnaire, it was decided to organize its analysis into three parts: i, analysis of the Portuguese citizens' perceptions about the exploitation of the metallic minerals in Portugal; ii, analysis of the Portuguese citizens' opinions regarding the importance of geoethics applied to

Table 1 Study sample characterization (n = 132)

Parameters analyzed		Frequency (f)	Percentage (%)
Gender	Female	56	42.4
	Male	76	57.6
Academic qualifications	Elementary education	4	3.1
	Secondary education	19	14.4
	Higher education	107	81.1
	Other (not specified)	2	1.5
Home area	North	87	65.9
	Center	23	17.4
	Lisbon (Capital)	12	9.1
	South (Alentejo region)	4	3.0
	South (Algarve region)	4	3.0
	Madeira's autonomous region	2	1.5
Professional situation	Student	46	34.8
	Employee	85	64.4
	Unemployed	1	0.8

mining in Portugal; and iii, analysis of the significance of the results. Then, the results are presented, using graphical analysis, elaborated for this purpose.

5.1 Portuguese Citizens' Perception of the Metallic Minerals' Exploitation in Portugal

This first section includes the results of questions one (Q1) to five (Q5), which focus on aspects of a scientific nature related to the concept of mineral (Q1), identification of metallic minerals (Q2), and metallic minerals that have lithium in their constitution (Q3), applications of lithium in society (Q4), and lithium exploitation sites in Portugal (Q5).

Regarding Q1, it appears that most participants selected two correct options (n = 50; 37.9%), and 39 participants (29.5%) selected the three correct options (Fig. 3). It is important to perceive that the most frequent responses, indicated by the participants, were: (C) It has a well-defined chemical composition and a crystalline structure (n = 98; 74.2%); (E) they are part of the constitution of all types of rocks (n = 84; 63.6%); and (D) they are solid and natural substances (n = 73; 55.3%).

In Q2, participants were asked to select three metallic minerals. Once again, the majority selected two correct options (n = 83; 62.9%) (Fig. 4). In this case, only 18 participants (13.6%) selected the three metallic minerals on the list provided (spodumene, wolframite, and magnetite). The metallic minerals most frequently selected by the participants are magnetite (n = 114; 86.4%) and wolframite (n = 97; 73.5%).

Fig. 3 Analysis of responses in Q1 (n = 132)

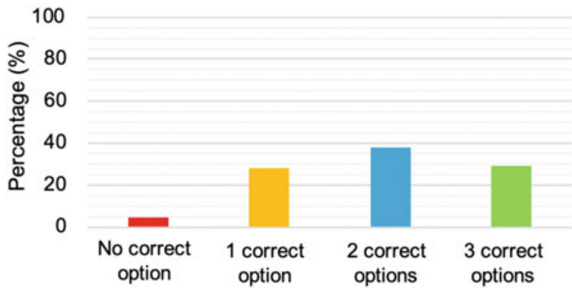
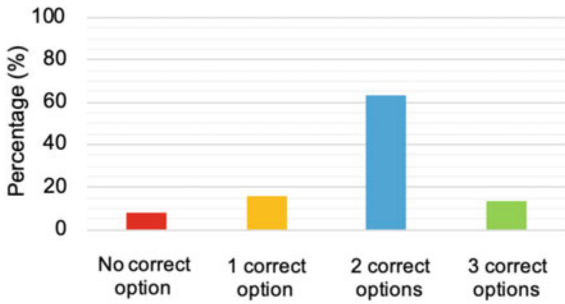


Fig. 4 Analysis of responses in Q2 (n = 132)

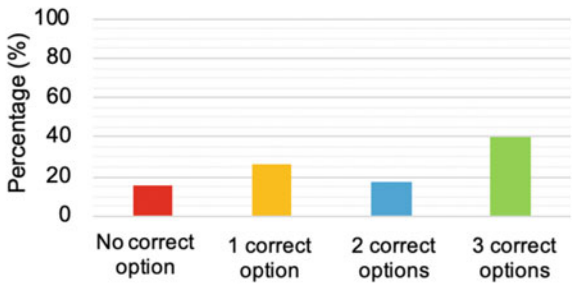


According to Fig. 5, in Q3, most participants selected three correct options (n = 53; 40.2%), that was three minerals that have lithium in their constitution, being the most referred: lepidolite (n = 95; 72.0%), petalite (n = 79; 59.8%) and espodumena (n = 67; 50.8%), which also correspond to the three correct options.

In Q4 (Fig. 6), about the applications of lithium in society, most participants selected two correct options (n = 91; 68.9%), and 36 participants picked the three correct options (27.3%). In this case, the most frequent options are the production of batteries for electronic devices (n = 128; 97.0%) and the production of batteries for electric cars (n = 122; 92.4%). Only 43 participants (32.6%) refer to ceramics as the use of lithium.

Finally, in Q5 (Fig. 7), there is a greater distribution of response categories, with a greater number of participants who did not select any correct option (n = 35;

Fig. 5 Analysis of responses in Q3 (n = 132)



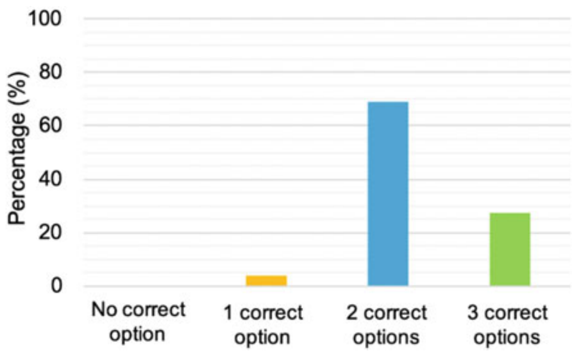


Fig. 6 Analysis of responses in Q4 (n = 132)

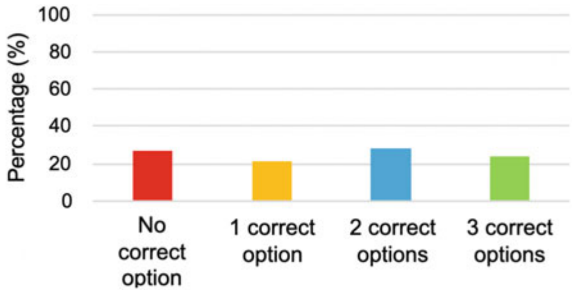


Fig. 7 Analysis of responses in Q5 (n = 132)

26.5%). Still, 37 participants (28%) selected two correct answers, and 32 (24.2%) selected three correct options.

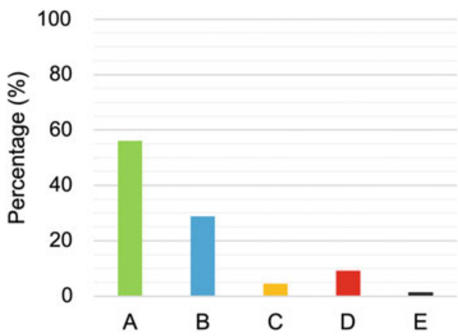
The most frequently mentioned lithium exploitation sites (in Portugal) are Covas do Barroso (Vila Real) (n = 79; 59.8%), Sepeda (Montalegre) (n = 67; 50.8%), and Gonçalo (Guarda) (n = 59; 44.7%) which, again, correspond to the three correct answer options.

5.2 Portuguese Citizens' Opinions Regarding the Importance of Geoethics Applied to Mining in Portugal

In this section, the geoethics related questions will be analyzed: firstly, the citizens' perceptions of its definition (Q6), its fundamental values (Q7), and then their opinions regarding its importance on reflection about mining in Portugal (Q8 and Q9).

Regarding the geoethics' concept (Fig. 8), it is possible to verify that most participants (n = 74; 56.1%) selected option A as the option that most closely matches the geoethics' definition known by them. Several participants (n = 38; 28.8%) selected option B, which, although similar to option A, has some details that make it less accurate to the geoethics' concept, for example, the fact that considering it as a science and for referring only to ethics (ethical values), not considering the other two values (cultural and social).

Fig. 8 Analysis of responses in Q6 (n = 132)



In this question, the participants were given the possibility of not agreeing with any of the options presented, enabling them to indicate their definition of geoethics, signing the option "other." There were 2 participants (1.5%) who selected this option. One answered, *"I don't know what STSE (Science, Technology, Society and Environment) means. It analyzes human intervention in the environment, finding sustainable solutions so that there is harmony between human intervention and the environment,"* and the other answered, *"Personally, it would mean sustainable exploitation or something similar."* The second answer has no relation to the definition of geoethics. Still, the first answer can be partially accepted because it reflects the effects human activities have on the environment. However, it does not refer to the consequences of human actions on the lithosphere.

In Q7, participants had to identify, for each statement presented, what is the inherent geoethical value: ethical value, cultural value, or social value. Figure 9 shows the percentage of participants who selected each of the three values in each statement (from 7.1 to 7.10).

It is important to mention that, in almost all statements, most participants correctly identified the geoethical value, corresponding to the highest percentage bar, to the geoethical value reflected in that statement. For example, statement 7.1. It concerned cultural values, as reflected in the yellow bar, indicating that most

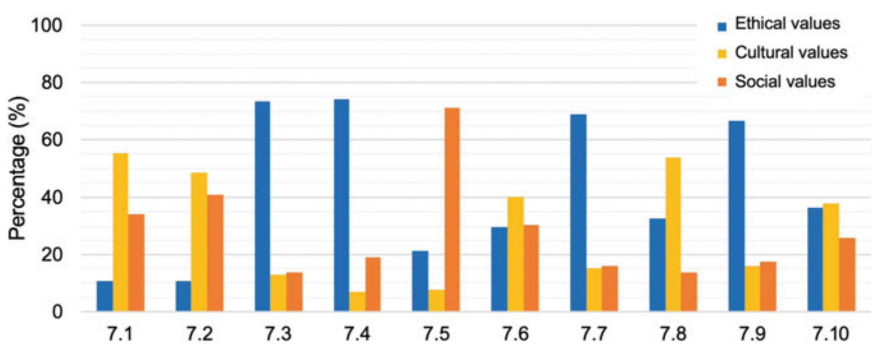


Fig. 9 Percentage of responses in each of the statements (from 7.1. To 7.10) of Q7 (n = 132)

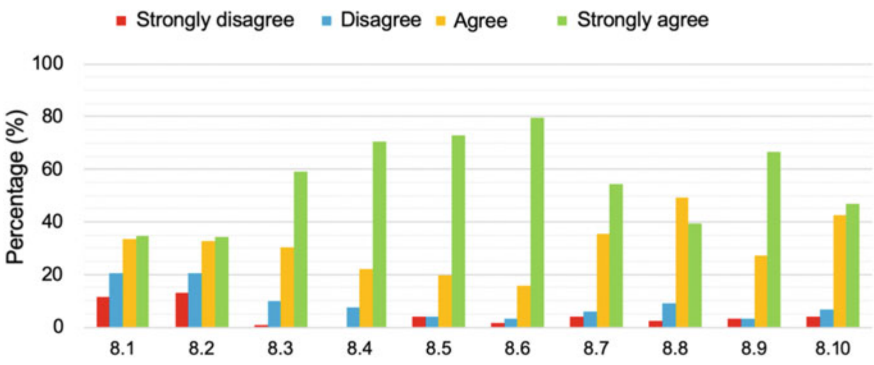


Fig. 10 Percentage of responses in each of the statements (from 8.1. To 8.10) of Q8 (n = 132)

participants (n = 73; 55.3%) selected this option. However, in statements 7.2 and 7.6, most participants selected options that do not correspond to the geoethical values reflected in the statements. Statement 7.2—they are concerned with developing citizens' scientific literacy concerning geosciences in general, concerns the social values of Geoethics, since scientific literacy is related to citizens' education/knowledge. Most participants considered it to be cultural values (n = 64; 48.5%).

In statement 7.6—they are concerned with geosciences education, for greater awareness and accountability of human beings' intimate relationship with the geosphere, social values were reflected again. In this case, most participants considered it to be cultural values (n = 53; 40.2%).

With questions Q8 and Q9, it was intended to analyze the opinions of citizens regarding the importance of geoethical reflection, firstly, in different situations (Q8) and, later, to behaviors/attitudes based on geoethical values, which must be considered when mining, namely lithium mining in Portugal (Q9).

Figure 10 refers to the percentage of responses from participants in Q8. It can be observed that in most situations (8.3, 8.4, 8.5, 8.6, 8.7, 8.9), the degree of complete agreement is visible by the green bar values, which appear to be clear. These situations concern: 8.3—Organization of geological hikes/tours accompanied by specialized guides in areas of geological and landscape interest (n = 78; 59.1%); 8.4—Construction of sustainable, self-sufficient, and environmentally friendly housing and commercial infrastructures (n = 93; 70.5%); 8.5—Studies for the location of safe zones for depositing radioactive waste from nuclear power plants (n = 96; 72.7%); 8.6—Environmental impact studies for the creation of quarries and mines (n = 105; 79.5%); 8.7—Construction of geoparks in areas of geological, environmental and cultural interest (n = 72; 54.5%); and 8.9—Actions to raise society's awareness of the importance of recycling and waste separation (n = 88; 66.7%).

In some statements, there was a greater distribution of responses/opinions. This is the case of statements 8.1 and 8.2, respectively, lithium exploration in rural areas (8.1) and offshore oil exploration and extraction (8.2). These are the statements with the highest percentages of disagreement (n = 27; 20.5% and n = 27; 20.5%, respectively) or total disagreement (n = 15; 11.4% and n = 17; 12.9%) of the participants.

In statements 8.8 and 8.10, there is some distribution of responses between the categories "agree" and "completely agree" (Fig. 10). However, it can be considered that, in general, the participants agree that a geoethical reflection applied in situations of integration of the elements of the extractive industry with other economic and social activities developed in the same region is important (8.8) and in conducting a geological drill in rock masses for analysis of radioactive element indexes (8.10).

Finally, in Q9, it appears that in almost all situations (except in 9.6), more than half of the participants consider the adoption of behaviors/attitudes based on geoethical values "extremely important" (Fig. 11).

Such values include: the professionalism and companies' transparency in revealing and explaining the type of intervention to be carried out with society (n = 97; 73.5%); the respect, consideration and attention of companies to the opinions of the population (n = 81; 61.4%); the understanding, respect, and trust of citizens for the work developed by companies (n = 69; 52.3%); clarity in communication between companies and citizens/municipalities and vice-versa (n = 94; 71.2%); the awareness of companies of the intimate relationship/dependence of human beings with the geosphere, concerned with the application of techniques that do not compromise their balance (n = 101; 76.5%); environmental concern, rethinking measures that do not have a negative impact on the environment (n = 94; 71.2%); innovation and implementation, within companies, of effective environmental quality assurance strategies (n = 89; 67.4%); ensure the control and minimization of waste production and harmful emissions, reducing practices that can affect the quality of life of future generations (n = 96; 72.7%); and the creation of clear and transparent laws and regulations, updated and adapted to the most advanced international levels of exploitation of mineral resources (n = 93; 70.5%).

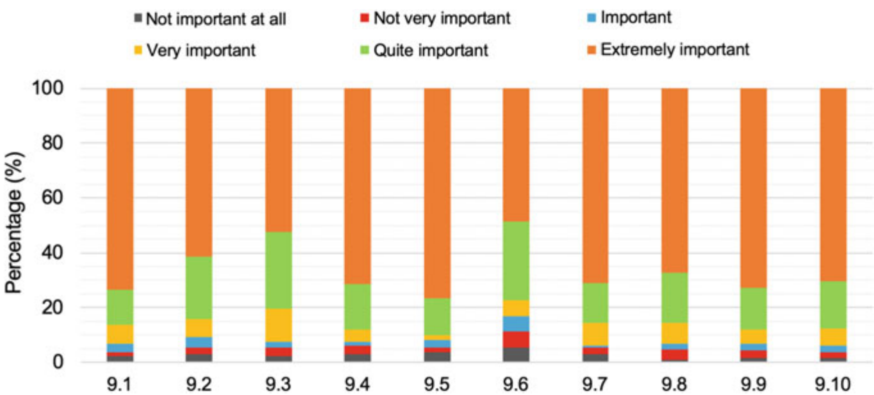


Fig. 11 Percentage of responses in each of the statements (from 9.1. to 9.10) of Q9 (n = 132)

5.3 Significance Analysis of the Questionnaire Results

The descriptive statistics applied to the data collected were considered relevant to ascertain any relationship between the variables studied. For this, the chi-square independence test was applied. This test allows us to verify if the distributions of two or more unrelated samples differ significantly concerning a given variable, whether they are independent of each other. The chi-square test was applied to the study variables, with some significant differences regarding the region of residence and gender. The analysis is presented below.

Regarding the region of residence, there are significant differences in the results of questions Q2, Q3, and Q4. Question 2 (Q2) concerns the identification of metallic minerals, and their distribution according to the home area is shown in Table 2. The chi-square test found that the differences in the responses in Q2, depending on the home area, are significant (χ^2 = 25.405; p = 0.05). In Q3, participants had to identify three minerals with lithium in their composition, also obtaining some differences in the distribution of results depending on the home area (Table 3). Also, in this case, the chi-square test determines that the differences in responses are significant (χ^2 = 16.574; p = 0.05). Regarding lithium applications

Table 2 Results in Q2, according to the home area (n = 132)

Home area	No correct option (*f*)	One correct option (*f*)	Two correct options (*f*)	Two correct options (*f*)
North	6	13	56	12
Center	2	5	16	0
Lisbon (Capital)	0	1	8	3
South (Alentejo region)	0	1	1	2
South (Algarve region)	2	1	1	0
Madeira's autonomous region	0	0	1	1
Total	10	21	83	18

Table 3 Results in Q3, according to the home area (n = 132)

Home area	Right answer (*f*)	Wrong answer (*f*)	Total
North	39	48	87
Center	5	18	23
Lisbon (Capital)	4	8	12
South (Alentejo region)	3	1	4
South (Algarve region)	0	4	4
Madeira's autonomous region	2	0	2
Total	53	79	132

Table 4 Results in Q4, according to the home area (n = 132)

Home area	Right answer (f)	Wrong answer (f)	Total
North	25	62	87
Center	4	19	23
Lisbon (Capital)	1	11	12
South (Alentejo region)	3	1	4
South (Algarve region)	0	4	4
Madeira's autonomous region	2	0	2
Total	35	97	132

Table 5 Results in Q6, according to the gender of the participants (n = 132)

Gender	Right answer (f)	Wrong answer (f)	Total
Female	25	31	56
Male	49	27	76
Total	74	58	132

(Q4), Table 4 shows the answers to this question, depending on the residence region. After performing the chi-square independence test, it was found that the differences in the responses in Q4, depending on the region of residence, are significant (χ^2 = 15.051; p = 0.05). In question Q6, on the definition of geoethics, significant differences were also found in the responses (Table 5), depending on the gender of the participants. Regarding the differences in responses in Q6 between female and male participants, the chi-square test allowed to verify that these differences are significant (χ^2 = 5.148; p = 0.05).

6 Conclusion

The work permitted us to realize that most participants know what a mineral is; identify some metallic minerals correctly and that they have lithium in their constitution; recognize lithium applications, at least the most publicized and discussed today, such as the production of electric car batteries and electronic devices; and identify some of the places in Portugal, where lithium exploitation is possible.

On the other hand, with this work, the importance of geoethics is highlighted mainly in the reflection on human activities that intersect the geosphere. This geoethical dimension must be the object of reflection, enabling to clarify the citizens on the geoethical concerns inherent to various human activities, namely the exploration of lithium, and other metallic minerals, in Portugal. The participants consider geoethical reflection relevant, whether applied to situations that reflect activities with some impact on the geosphere (e.g., development of geoparks in areas of geological interest; creation of quarries and mines) and appraise geoethical reflection relevantly when related to essential attitudes and behaviors in activities related to mining (e.g., professionalism, respect, and communication between the exploration and exploration companies).

We could finish this work, questioning if it will be worth exploring the lithium in Portugal or whether it will be preferable to think only of preserving the environment. However, it would be undeniable that many more questions would have to be added to this question. Notably, whether in the age of technology, we would live without the exploitation of metallic minerals and whether it would be possible to alienate ourselves from gadgets and start having an ecocentric relationship with nature. Plagiarizing a Portuguese radio host, we chose to end with this simple reflection: Is it worth thinking about it.

Core Messages

- Modern societies need raw materials, which implies the need to establish good public communication related to mining.
- Mining damages are about the land and environment, but also about the profound ecological and cultural importance that the sites have to people who have lived there for many years.
- The teaching and dissemination of Geoethics values are essential at schools, universities, and science communication areas.

Appendix I: Questionnaire on the Use of Metallic Minerals in Portugal

Gender: M ____ F ____

Age (until 31st December 2019): _______

District where you live: _______________________________________

Education: ____

Job: ____

(In each of the questions 1 through 6, the correct answers are underlined)

1. From the following options, select the one(s) that corresponds to the characteristics of the minerals:

A. Natural or synthetic substance.
B. It is obtained from the alteration of the rocks to the Earth's surface.
C. It has well-defined chemical composition and crystalline structure.
D. They are solid and natural substances.
E. They are solid and natural substances.
F. They can be obtained in the laboratory through physicochemical reactions similar to those that occur in nature.

2. From the following mineral listing, select three metallic minerals:

Quartz	Staurolite	Moscovite
Spodumene	Calcite	
Lead	Rock salt	Albite
Volframite	Biotite	
Magnetite	Halite	

3. From the following listing of metallic minerals, select three minerals from which lithium is extracted:

Petalite	Cassiterite	Calcopyrite
	Spodumene	Lepidolite
Magnetite	Pyrite	
Volframite	Galena	

4. From the following listing, indicate three uses of lithium:

Production of electric car batteries.
Production of batteries of electronic devices.
Obtaining metal alloys.
Production of electrical cables.
Jewellery.
Ceramic.

5. For ore exploitation sites in Portugal, select three where lithium extraction occurs:

Covas do Barroso (Vila Real)	Sepeda (Montalegre)
Panasqueira (Fundão)	Moncorvo (Bragança)
Neves-Corvo (Beja)	Aljustrel (Beja)
São Pedro da Cova (Gondomar)	Montemor-o-Novo (Évora)
Gonçalo (Guarda)	

6. Considering the following statements, select the one that is closest to the concept of geoethics.

a) Branch of knowledge that reflects on the basic values of human behaviors and practices, when their activities intercept the geosphere.
b) Science responsible for the study of ethics applied to planet Earth and all practices that affect its balance.
c) Area of knowledge that analyzes human practices and their impact on the environment, in order to monetize them.
d) Science responsible for the analysis of human practices from a STSE perspective, with a view to sustainable development.
e) Other ____

7. For each of the following statements, select the geoethical value that concerns you *(Correct answers marked with X)*

Statements	Ethical values	Cultural values	Social values
1. Influence society's current and future perceptions of space and time, especially in Western cultures		X	
2. Are concerned with the development of citizens' scientific literacy to geosciences in general			X
3. Understand the respect for natural systems and their dynamics in any intervention in the environment	X		
4. Relate to the individual sphere of each geoscientist and the integrity with which he develops his work	X		
5. Relate to the well-being of people and the measures necessary for their survival and protection			X
6. Are concerned with geosciences education for greater awareness and accountability of the intimate relationship of human beings with the geosphere			X
7. Comprise high professionalism and competence, to guarantee the reliability of the studies and the transparency of the results	X		
8. Relate to the appreciation of the geological heritage, including its aesthetic value and its impact on the relationship between humans and the geosphere		X	
9. Allow the geoscientist to reflect on his work, giving it a more profound meaning	X		
10. Allow the recognition of geodiversity, contributing to a more effective defense against natural risks		X	

8. The following statement **Importance scale** s refer to situations in which geoethical reflection may be necessary. Indicate, for each statement, the degree of agreement, on a scale of 1–4 (1—strongly disagree; 2—disagree; 3—agree; 4—strongly agree).

Statements	Concordance			
	1	2	3	4
1. Lithium exploitation in rural areas				
2. Offshore oil exploration and extraction				
3. Organization of geological hikes/tours accompanied by guides specialized in areas of geological and landscape interest				
4. Construção de infraestruturas habitacionais e comerciais sustentáveis, autossuficientes e amigas do ambiente				
5. Estudos para localização de zonas seguras para depósito de resíduos radioativos de centrais nucleares				
6. Estudos de impacto ambiental para criação de pedreiras e minas				
7. Construction of sustainable, self-sufficient, and environmentally friendly housing and commercial infrastructures				
8. Integration of extractive industry elements with other economic and social activities developed in the same region				
9. Activities to raise society's awareness of the importance of recycling and waste separation				
10. Conducting drilling work in rock masses to analyze radioactive element indexes. Conducting drilling work in rock masses to analyze radioactive element indexes				

9. The following statements refer to behaviors/attitudes related to mining (namely, lithium mining in Portugal). Considering the geoethical principles, indicate the importance attached to each of the practices mentioned, on a scale of 1–6 (1—nothing important; 6—extremely important).

Statements	Importance scale					
	1	2	3	4	5	6
1. Professionalism and transparency of the exploration companies in disclosing and explaining the type of intervention to be carried out, among society						
2. Respect, consideration, and attention of companies to the population's opinions and concerns in the intervention area						
3. Understanding, respect, and trust of the citizens for the work developed by the technicians and specialists of the exploration companies						

(continued)

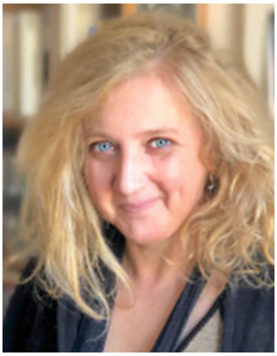
(continued)

Statements	Importance scale					
	1	2	3	4	5	6
4. Clarity in the communication between the exploration companies and the citizens/municipalities and vice-versa						
5. Awareness of exploration companies about the intimate relationship/dependence of human beings with the geosphere, concerned with the application of techniques that do not compromise the balance of this system						
6. Cultural implications of mining in the geodiversity, including changes in the natural landscape and geological heritage of the region under intervention						
7. Environmental concern and redefinition measures that do not have a negative impact on the environment						
8. Innovation and implementation, within companies, of effective environmental quality assurance strategies						
9. Ensure the control and minimization of waste products and harmful emissions, reducing all practices that may compromise the quality of life for future generations						
10. Creation of legislation, clear, transparent, updated and adapted to the most advanced international levels of exploitation of mineral resources						

References

1. Ciência Viva (2019) 150 anos da Tabela Periódica dos Elementos Químicos. Ciência Viva—ANCCT. Available via http://www.cienciaviva.pt/tabelaperiodica. Accessed 21 Jul 2019
2. Peppoloni S, Di Capua G (2015) The meaning of Geoethics. Geoethics: ethical challenges and case studies in Earth Science, 1st edn. Elsevier, Massachusetts. doi:https://doi.org/10.1016/B978-0-12-799935-7.00001-0
3. Lucchesi S, Giardino M (2012) The role of geoscientists in human progress. Ann Geophys 55 (3):355–359
4. International Association for Promoting Geoethics (2019) INGV. IAPG. Available via https://www.geoethics.org/. Accessed 21 Jul 2019
5. Vasconcelos C, Torres J, Vasconcelos L, Moutinho S (2016) Sustainable development and its connection to teaching geoethics. Episodes 39(3):509–517
6. Pievani T (2012) Geoethics and philosophy of Earth sciences: the role of geophysical factors in human evolution. Ann Geophys 55(3):349–353
7. Matteucci R, Gosso G, Peppoloni S, Piacente S, Wasowski J (2012) A Hippocratic Oath for geologists? Ann Geophys 55(3):365–369
8. Lugeri F. R., Farabollini P., Graziano G. V., Peppoloni S. (2013) Geology and society: new perspectives. In: Abstracts of IX Forum di Scienze della Terra. Pisa 15–18 Sep 2013
9. Peppoloni S, Di Capua G (2017) Geoethics: ethical, social, and cultural implications in geosciences. Ann Geophys 60:1–8
10. Arvanitidis N, Boon J, Nurmi P, Di Capua G (2017) White paper on responsible mining. In: International Association for Promoting Geoethics. APG. Available via http://www.geoethics.org/wp-responsible-mining. Accessed 21 Jul 2019

11. Boon J (2020) Relationships and the course of social events during mineral exploration—an applied sociology approach. Springer Briefs in Geoethics, vol 1, 1st edn. Springer International Publishing
12. Christmann P, Gloaguen E, Labb JF, Melleton J, Piantone P (2015) In: Alexandre Chagnes, Jolanta Swiatowska (eds) Global lithium resources and sustainability issues. Global lithium process chemistry: resources, extraction, batteries and recycling, 1st edn. Elsevier, Amsterdam, pp 1–40
13. Viegas H, Martins L, Oliveira D (2012) Alguns aspectos da geoestratégia global do lítio O caso de Portugal. Geonovas 25:19–25
14. Young W (2009) Review of lithium effects on brain and blood. Cell Transplant 18:951–975
15. Carvalho JMF, Farinha JALB (2004) Lithium potentialities in Northern Portugal. In: Editor 17th industrial minerals international congress, Barcelona 28–31 Mar 2004
16. Lima A (2019) Lítio, um metal para o futuro de Portugal? Público. Available via https://www.publico.pt/2019/05/13/economia/opiniao/litio-metal-futuro-portugal-1872284. Accessed 21 Jul 2019



Clara Vasconcelos, who graduated in Geology and with Ph.D. in Education, is an Associate Professor with Habilitation in the Department of Geosciences, Environment, and Spatial Planning, and in the Science Education Unit of the Faculty of Science of Porto University. She researches Science Education and Teaching Methods, Sustainable Development, and Geoethics at CIIMAR. Clara is Commissioner of IUGS in COGE (2018-), National Co-Coordinator of International Association for Promoting Geoethics (IAPG)—Portugal (2015-), Member of the Panel of Experts of Education of the European Federation of Geology (2014-), and Portuguese Council Member of the IGEO (2010-).

Alexandre Lima holds a Ph.D. degree (2000) in geology at the University of Porto, Portugal, in co-tutelle with INPL, Nancy, France. Since 2000, he is an Associate Professor at the University of Porto. He has experience working in exploration, geochemistry, and mineralogy of rare-metal and industrial mineral deposits and more than 15 years' post-doctoral research experience in the exploration, nature, and origin of pegmatites and related mineralization. His research has led to the publication of 36 papers in peer-reviewed journals in earth sciences. For three years, starting in 2014, he was responsible in Portugal for the project SUSMIN *"Tools for sustainable gold mining in EU."* In 2015 he was a Researcher in the ERA-MIN project NEWORES *"Development of New models for the genesis of Rare Metal (W, Nb, Ta, Li) Ore deposits from the European Variscan Belt."* In 2018, he started the LIGHTS *"Lightweight Integrated Ground and Airborne Hyperspectral Topological Solution."* Recently, in 2020, the GREENPEG "New Exploration Tools for European Pegmatite Green-Tech Resources." Alexandre is responsible in Portugal for both projects dedicated to Lithium exploration tools.

18 Creative Approaches for Mitigating Environmental Challenges: A Cosmopolitan Localism Perspective

Ginn Bonsu Assibey and Alettia Vorster Chisin

We have forgotten how to be good guests
How to walk lightly on the earth as its other creatures do.

Barbara Ward

Summary

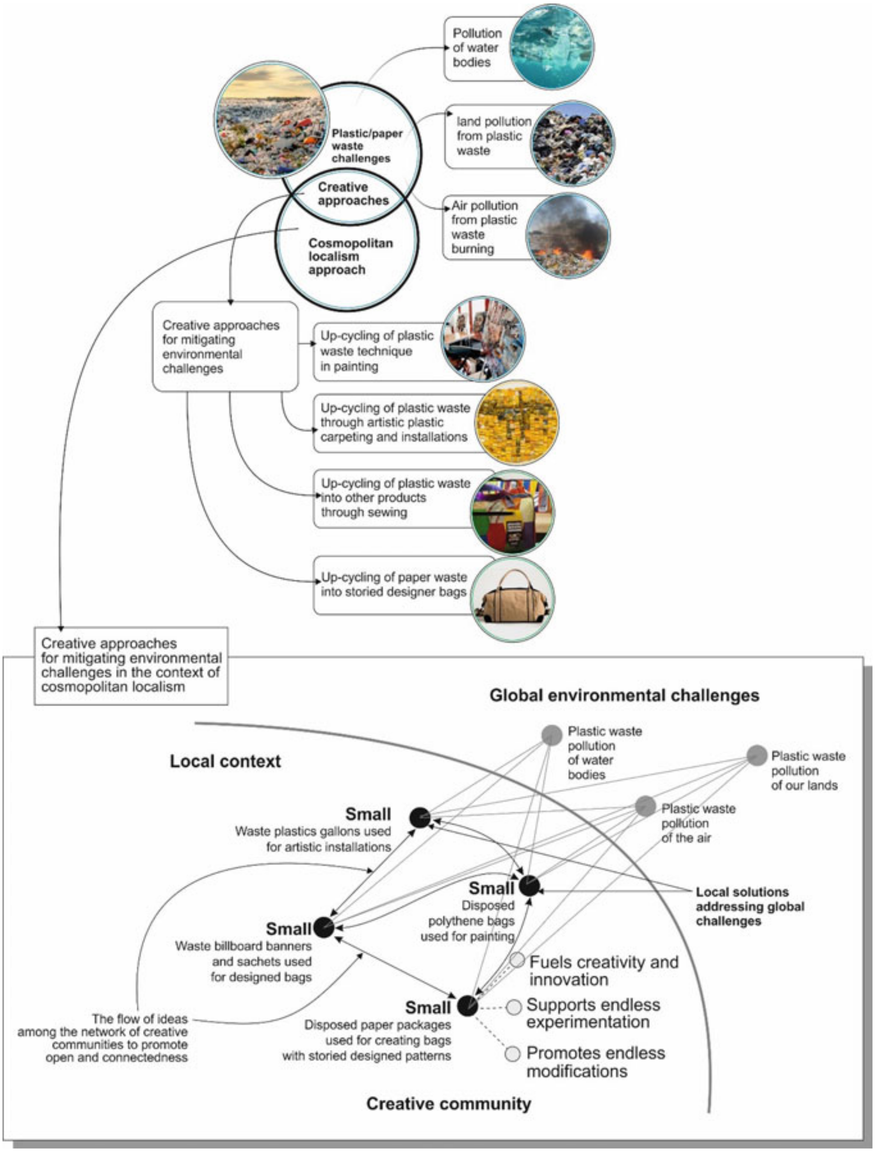
The detachment of economic growth from environmental challenges has been an active, though challenging, area of endeavor. Various interventions have taken into account the problem involving policymakers in technological and social systems. However, little improvement has occurred. Therefore, this chapter suggests that one approach to tackle this is to promote cosmopolitan localism from the design and fine art perspectives to reduce environmental challenges. Artists and designers are creative and can leverage these creative abilities in many ways. Thus, we further discuss the artists' and designers' role in advancing cosmopolitan localism by showing emerging creative scenarios centered on the recycling and up-cycling of waste materials present in the environment into designed artifacts. These scenarios also exhibit cosmopolitan localism in small, local, open, and connected fibers to promote a sustainable environment.

G. B. Assibey (✉) · A. V. Chisin
Department of Applied Design, Faculty of Informatics and Design,
Cape Peninsula University of Technology, Cape Town, South Africa
e-mail: 214329453@mycput.ac.za

A. V. Chisin
e-mail: chisina@cput.ac.za

G. B. Assibey
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Cape Town, Ghana

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_18



Creative scenarios for mitigating environmental challenges in the context of cosmopolitan localism.

The code of this chapter is 01000011 01101111 01110011 01101101 01101111 01110000 01101111 01101100 01101001 01110100 01100001 01101110 00100000 01101100 01101111 01100011 01100001 01101100 01101001 01110011 01101101 00001101 00001010.

1 Introduction

The link between economic development and ecological challenges may seem weak. However, the crucial need for product protection from spoiling—partly to enhance profit optimization, puts under the spotlight one of the ubiquitous materials for product protection from decay, namely plastic packaging. The link becomes lucid through probing the materials utilized by designers. There are several traditional packaging materials obtained from natural resources used to manufacture carrier packages, which come with associated challenges resulting from single-use and disposal problems. Other challenges may come about due to the processing and related waste chemical substances, which in some instances flow into rivers and onto the ocean. Another challenge is caused by the fumes from the energy used to power the manufacturing machines that pollute the air. The waste substances discharged into water bodies and the atmosphere threaten the lives of the living organisms on the earth [1]. Apart from the by-products of materials' processing, one aspect of product waste that needs urgent attention is packaging waste due to its high volume. Moreover, it is dispensable and occupies a portion of the solid waste from industries. As other continents endeavor to upscale their economies amid an increase in their population, package production will also shoot up, consequently yielding more waste [2].

Some of these waste materials end up in rivers, streams, and gutters, which poison and cause flooding due to drainage blockage [2]. From the yearly packaging growth (Fig. 1), it is easy to understand that packaging volumes will increase when economic and population growth are factored in the time ahead. Recycling has, therefore, become the exclusive approach to using waste as practiced by western countries. However, some "useful" waste goes untapped, especially in Africa, due to limited recycling infrastructure. In support of the limited infrastructure, The Guardian has disclosed that only 14% of plastic package waste is recycled even on a worldwide scale. Thus, it triggers accountability questions for the rest of the 86% [5]. Though packaging serves as a panacea in advancing growth in the food industry, the end-users are not knowledgeable about its post-usage management [3].

The challenge now is how designers and artists can leverage their creative potential for up-cycling of used packages or other natural materials in different ways to reduce and manage waste. Baldacchino and Cutajar advance an urgent need to shift to a culture that nurtures sustainability to keep the earth's productive system intact, to support current lives and future generations [6]. Fortunately, many stories of artists' environmental sustainability initiatives are in motion and tagged as recycled art and design. It implies that we do not have to start from zero or reinvent the wheel—but we do need to spread the stories to activate the creative thirst for up-cycling and repurposing of plastic or paper waste into artifacts of worth. Therefore, this chapter gives narrations on the various recycled or up-cycled arts and designed products emerging on the African continent. We also advance that the recycled arts and designed products are happening within the context of cosmopolitan localism, which serves as a platform for artists to innovatively and

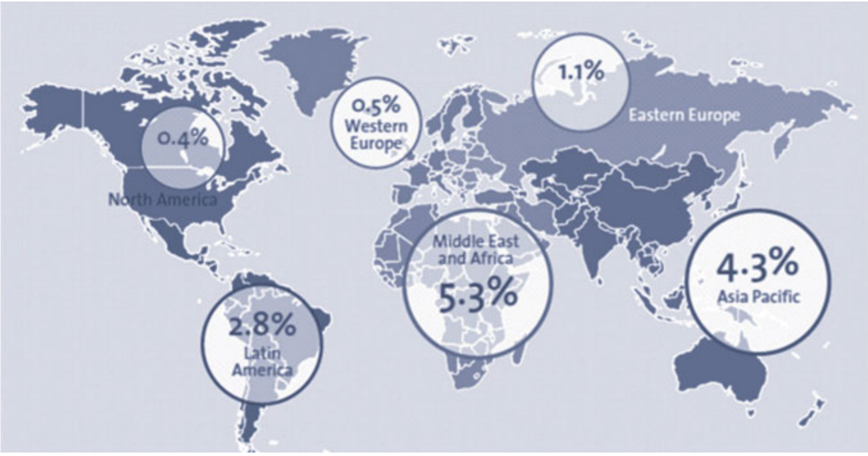


Fig. 1 Forecast of annual packaging volume growth rates (2014–2019) (adapted from [14])

creatively engage in experimenting with different waste materials to understand how to use them. In doing so, a hub can be created to share their innovations for environmental sustainability in the creative arts and design—to promote these practices to become the norm of the day.

2 Cosmopolitan Localism

In the context of environmental sustainability, there are different global models available but limited in practice due to several problems linked to applying a global solution to a local challenge [7]. Several global solutions might not be fruitful when used locally without contextualizing them. Thus, we need to reroute our path to become more resilient to emerging social and economic challenges through deviating from ingrained paths in innovations and creativity [7]. Instead, we must embrace cosmopolitan localism and employ and practice its global ideas through local approaches [7]. Embracing cosmopolitan localism will present several challenges, but we need not reinvent the wheel because there are evolving *concrete experiences that could consolidate and spread to become the most convincing answers to the dramatic challenges that we must now begin to face* [8].

Cosmopolitan localism is based on four main qualities: small, local, open, and connected. Small in the context of cosmopolitan localism connotes linked creative communities that solve similar challenges. Thus due to the linked creative communities, the small concept becomes enlarged when viewed from a connected perspective. In this situation, the small creative communities are easy to manage and control, and that permits the application of global ideas in localized ways. Manzini advances that these are qualities of cosmopolitan localism derived from *"synthesizing the results of 20 years of discussions and concrete experiences, which indicate that there is no hope for designing sustainable solutions without starting from the notions of local and of the community to which this local mainly refers"* [8]. Based on his experiences, Manzini advocates that sustainability will only thrive through solutions that factor local approaches in our contemporary dispensation [8]. The need to embrace cosmopolitan localism to advance environmental sustainability serves as the platform for this chapter. We advance that there are already emerging solutions or scenarios that need to be explored and diffused to promote cosmopolitan localism. Therefore, the chapter brings to bear how designers and artists are harnessing their creative abilities to mitigate environmental challenges to change the status quo.

3 Emerging Scenarios for Mitigating Environmental Challenges

Plastic and paper waste materials have been a menace for ages. Some countries have banned plastic use due to its impact on the environment and other living organisms. However, some countries have opted for a flexible approach by educating citizens regarding the issue, while others do not have any strategy. The result is that plastic wastes still occur, littering our environment, choking gutters and water systems, leading to floods, and finally filling the ocean. Some countries have adopted recycling and reuse to mitigate the environmental challenges caused by plastic waste disposal. Among the disciplines that impact plastic waste's up-cycling, among other waste materials, are artistic painting and fashion design in Africa. The following sections give four case details on waste materials up-cycling techniques in South Africa and Ghana.

3.1 Up-Cycling of Plastic Waste Through Artistic Painting

Up-cycling in this chapter comes from cosmopolitan localism centered on advancing a worldwide concept through local means. Thus, the idea of environmental sustainability and its associated mainstream approaches are not cast in stone. Other contextualized approaches are emerging to mirror the same environmental sustainability concept using off-the-grid approaches in Africa. One such approach is the up-cycling of waste materials through creativity. Up-cycling is an approach that sits between upgrading and recycling materials [9]. Up-cycling means creating something new from waste or old materials without any physical transformation of the material. Let us now look at the up-cycling approach used by a painter in South Africa—Mbongeni Buthelezi. Mbongeni Buthelezi, a prominent plastic artist, summed up his technique with the phrase, *"I collect rubbish and create something beautiful from it. I collect something that has no value and give it new life…"* [10]. He collects all sorts of plastic waste and creates a plastic waste bank, as shown in

Fig. 2 Waste plastics gathered by Mbongeni Buthelezi to be used for his plastic collage (adapted from www.pinterest.ch)

Fig. 2, from which he fetches his plastic for his artistic expression. During the collection process, he selects the appropriate plastics he can work with easily.

3.2 Up-Melting of Plastic Waste in Painting: Plastic Collage

Let us now expand on the up-cycling technique by Mbongeni Buthelezi for his plastic impressions. He describes his style as a plastic collage. He first cuts the plastics into strips, then prepares his wooden stretcher, and covers it with several plastic roofing sheets. Using a watercolor approach, he skillfully melts the plastic strips with a heat gun onto the canvas prepared with the roofing plastic. Figure 3 shows Buthelezi at work. What interests him in using waste plastic for his artistic expression is the link to the plastic's "spontaneous" nature and the freshness it brings to the artistic interpretation because of its watercolor-like nature, once applied. This approach does not allow him to experiment "outside" of his mainstream artistic activities. Thus, every piece of plastic he sticks down by melting provides a learning experience. In essence, he develops his style as he works since

Fig. 3 Mbongeni Buthelezi at work (photo by Wanda Hennig [15])

he is a pioneer in plastic usage for artistic expression. He uses approximately five thousand (5000) pieces of plastic in each artwork he produces [10].

This discussion of Mbongeni Buthelezi's approach illustrates how up-cycling of waste plastics that hitherto were either burnt or carried away into water bodies to cause environmental damage turned into valuable art pieces. Apart from the plastic art creations, he travels abroad from time to time to share his knowledge with students and communities and to share his considerable skills at the local community level. Figure 4 shows a masterpiece of this plastic collage technique developed and refined by Buthelezi.

3.3 Up-Cutting of Plastic Waste Through Artistic Plastic Carpeting and Installations

Another artist involved in reducing plastic pollution is Serge Attukwei Clottey from Ghana. He uses gallon drums and has termed his artistic style as "Afrogallonism," an artistic, innovative use of the yellow gallon containers used as cooking oil canisters and then repurposed to collect water or fuel [11]. He cuts the gallons into pieces and then creates holes at the edges to facilitate sewing the pieces of plastics and a strong thread. He paints some of the plastic pieces based on the concept he wants the carpeted plastic pieces to convey. Figure 5 shows an artistic carpeted-plastic artwork installation by Serge Attukwei Clottey, while Fig. 6 represents an artistically sewn plastic carpet.

Fig. 4 Mbongeni Buthelezi's plastic collage sample works [16]

Fig. 5 Plastic carpet artwork installation by Serge Attukei Clottey (photo by Serge Attukei Clottey [11])

Fig. 6 Sown plastics forming an artistic carpet by Serge Attukei Clottey (photo by Serge Attukei Clottey [17])

3.4 Up-Sewing of Plastic Waste into Designer Products

This type of up-cycling is not only artistic but also represents an innovative approach. Among the companies engaged in up-cycling of plastic waste into products is Trashy Bags. Trashy Bags is a social-environmental enterprise based in Accra, Ghana, that up-cycles waste plastics into bags and souvenirs [12]. The designers collect, clean, and stitch plastic trash such as water and fruit juice sachets into esthetically pleasing and durable bags, among others [12]. The enterprise also makes bags from used advertising billboard banners, and designers have collected and up-cycled approximately twenty million plastic sachets since 2007 [12]. Every month about two hundred thousand (200,000) plastic sachets are collected and up-scaled into laptop bags, messenger bags, backpacks, and other everyday use bags [12].

The central motive for up-cycling is not only business-driven (profit) but also to aid in the reduction of plastic waste management in Ghana [12]. Only two percentage is recycled, while the ninety-eight percentage ends up polluting the environment and destroying water bodies [12]. These examples of up-cycling in Africa testify to the need for revamping innovative approaches to tackle the environmental challenges caused by plastic waste on the continent through up-cycling options. Through comparing the volume of the plastic waste generated against the percentage recycled, it is clear that the up-cycling strategies may process but a handful of the overall plastic waste generated. Thus, up-cycling can be considered an additional solution to support existing permanent recycling plans from waste management companies.

3.5 "Up-Storying" of Waste Paper into Durable Design-Storied Bags

A second sustainable environmentally inclined business from South Africa is called Wren Design. This company takes ordinary waste papers/paper packaging and reimagines them through folding, stitching, fusing, and coating with water-resistant materials into durable storied bags [13]. The storied bags' concepts are in the designed patterns of the bags, the construction methods, and the continuous product development, which make them unique. Thus, every design pattern on the bag carries a story. These storied bags are patented; however, the process of up-cycling the paper waste into durable materials is open and can be practiced by those who are interested. Figure 7 shows the storied paper waste bags designed by the founder of Wren Design.

Fig. 7 Storied waste paper designer bags [13]

4 Innovative Up-Cycling Strategies through the Lens of Cosmopolitan Localism: Toward Conceptualizing Up-Cycling Innovations for Environmental Sustainability

All four cases could show how innovative and environmentally conscious African artists and designers contribute their quota to reducing plastic and paper waste in two countries on the continent. These up-cycling strategies show alternative approaches to plastic and paper waste reduction, developed by artists and designers harnessing their unique creative abilities that embrace the concept of cosmopolitan localism. Up-cycling on its own is not the long-term solution to the environmental plastic waste challenge. In conjunction with organized, official measures, it functions as an additional measure to reduce plastic and paper waste pollution and serves as an income stream for some designers, artists, and ordinary people in Ghana and South Africa. Let us now look at these four cases in the context of *small, local, open, and connected,* which are the pillars of cosmopolitan localism.

4.1 Innovative Up-Cycling of Plastic and Paper Waste in the Context of Small

When four cases become perceived in the context of small, it manifests in the volume of artifacts produced innovatively and the working space and materials allocated for the creative—or the design work. It implies that designed artifacts occur on a small scale in small working spaces, not in factory settings. In cosmopolitan localism, small relates to fueling artistic and designerly innovations through experimentation and an open-ended perspective. Whatever artifact is produced in the small context, it may be seen to be an experimental piece that can be modified endlessly to suit a new context.

On the other hand, small is not small from a technological perspective or from disseminating information and marketing perspectives. All four cases discussed for their innovative approaches regarding the mitigation of plastic and paper waste are available on the internet. The idea behind the online presence of these innovative businesses is not only to market the products but also to educate the public about the possibility of engaging in similar creative work to generate income and foster environmental consciousness in sustainability. Thus, the internet has become the visual hub for connectivity among the actors facilitating cosmopolitan localism.

As other people in Ghana and South Africa engage in similar creative work toward environmental sustainability, the number of people practicing waste up-cycling in art and design will increase. Thus, the picture painted at the end of the day will be a network of small creative nodes with artists and designers working on similar projects. When all the innovation nodes within the local context come together, they will no longer be small, as advanced by Manzini [7]. The small concept also equips the actors in the nodes with the freedom to experiment, driven by the motivation to obtain (similar) desirable results, yet with room for endless modifications, as shown in Fig. 8. The nodes thus become creative communities with endless possibilities in terms of sustainable environmental approaches.

4.2 Innovative Up-Cycling of Plastic Waste in the Context of Local

The cosmopolitan localism concept regarding up-cycling in the four cases described unearths the uniqueness and diversity of creativity driven by the availability of local waste materials. The waste materials used in these four cases illustrate the kinds of waste material available within a locality. The innovative approach is leveraged for converting waste into artifacts of worth. Thus, the transformed products inherently bear the mark of the waste materials which were locally harvested. Though the artifact produced might have local tags, the concept of up-cycling has a global underpinning anchored to environmental sustainability and, in these cases, available to a global market to a greater or lesser extent. Therefore, the local essence is not necessarily only locally focused but globally “acted” in its availability. The four cases discussed show that the abundance of local materials triggers local ideas with

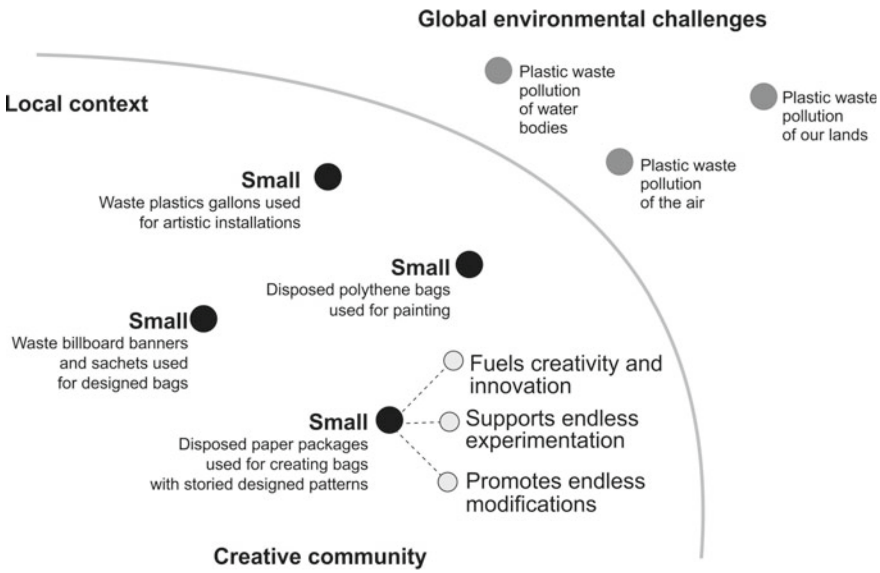


Fig. 8 "Small" manifestation and benefits in up-cycling of plastic and paper waste

international essence, which bridges the gap between local creativity and global connectivity.

Manzini (2010) advanced that in the context of localizing global concepts, the local ceases to be local anymore. However, we hold a more nuanced view. Using a local lens to address a global challenge or spearhead a global concept does not necessarily transform the local artifact into a global artifact; it will retain provenance. Thus, local remains local since the up-cycled product becomes unique due to the quality and creativity attached to a specific milieu. The local rather becomes a yarn woven into the global fabric of innovative approaches for advancing sustainable environmental practices. Therefore, the local innovative up-cycling approaches are recognized as part of the global cloth for addressing challenges in situationally fitting ways. In other words, global challenges emanate from different locales and require local solution-inclined approaches, as witnessed in the four enterprises discussed. In a nutshell, the concept of local up-cycling in the framework of cosmopolitan localism represents a global phenomenon adapted and applied contextually to have a global impact (even if limited) in addressing environmental challenges, as shown in Fig. 9.

4.3 Innovative Up-Cycling of Plastic Waste in the Context of Open and Connected

Open in the context of cosmopolitan localism signifies the creative communities' actors' open-mindedness and ability to welcome a global flow of ideas or concepts

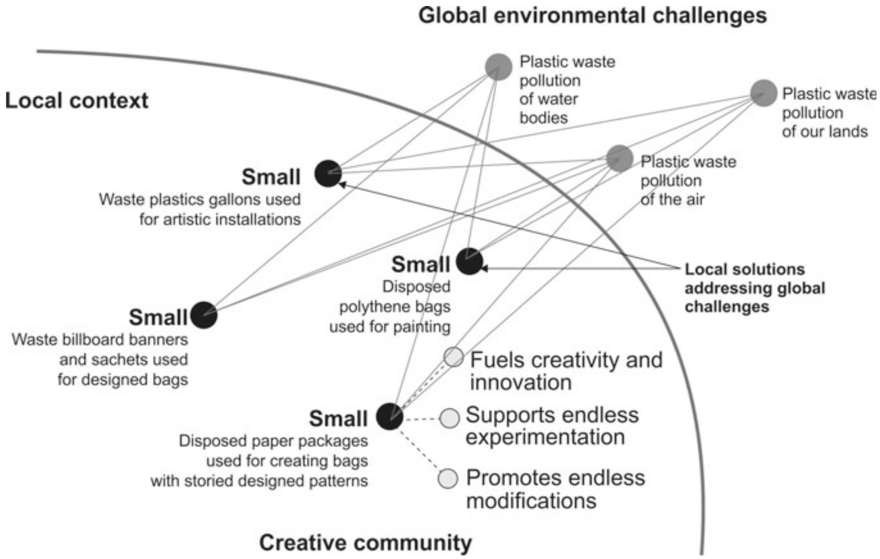


Fig. 9 Local solutions with global impact

into their locality. These are ideas practically implemented using indigenous materials, especially in the area of design. Localizing global ideas requires rethinking a specific perspective and developing home-grown solutions that reflect these ideas. The environmentally sensitive and innovative approaches discussed in this chapter illustrate the creative communities' open mindset, which contributes to making their innovations relevant in the world. This open mindset (which is asset-based) strengthens the locals' socio-technical systems and makes them dynamic, solution-oriented, and resilient in an era where sustainable challenges emerge in various forms.

The artists' and designers' flexible styles create room for connectedness among concurring local actors and those who learn from them through access to technology and the internet. The access to countless creative communities' innovations shared on the internet also creates an avenue for fueling other local actors' imaginations and innovations in birthing similar creative communities through design adaptation. Thus, technology in the form of the internet serves as a hub for triggering and fueling actors to expand creative communities' networks. When perceived on a larger scale, these up-cycling creative communities form a network focused on solutions toward environmental challenges, making them stronger and resilient, as captured in Fig. 10. These creative communities' innovative approaches position them as dynamic design instruments ready to tackle uncertainties in the future by harnessing their abilities for alternative solutions that are locally based.

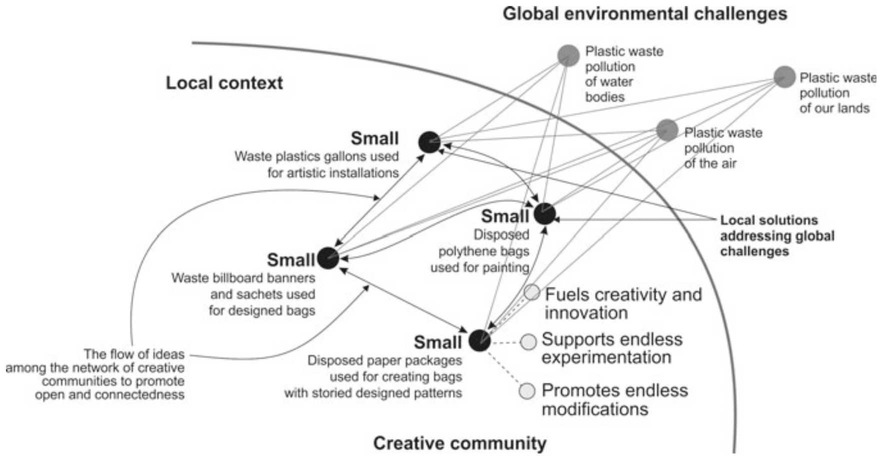


Fig. 10 The flow of ideas among the network of creative communities to promote openness and connectedness

5 Conclusion

It has been difficult, if not impossible, to detach plastic waste generation resulting from an economic activity from such activity. Economic development, manufacturing, and packaging seem to go hand in hand. However, the emerging creative and sustainable product scenarios discussed in South Africa and Ghana show hope for the future. Artistic up-cycling approaches to plastic waste fuel the hope of more artifacts of worth being derived from such waste. Driven by the framework of cosmopolitan localism, the up-cycled plastic-waste approaches resulting in desirable artwork and products, coupled with the creatives' vision, show traces of small, local, open, and connected fibers which promote environmental sustainability and make the local and global inseparable.

Core Messages

- Artists and designers need to explore the options available in local and valuable waste materials actively to contribute to the reduction of environmental challenges.
- Artists' and designers' websites must feature skills sharing to promote skills transfer worldwide for a collective plastic waste reduction approach.
- Artists' and designers' creative work and approaches should reduce global environmental degradation caused by waste, even if to a small degree.
- Creatives need to work collaboratively to establish a waste laboratory that incubates creative up-cycling techniques that can be disseminated worldwide to help them think locally and act globally.

References

1. Benson E, Perullo Y (2017) Design to renourish: sustainable graphic design in practice. CRC Press
2. Jindal M (2010) Unpacking the packaging: environmental impact of packaging wastes. J Environ Res Dev 4:1084–1092
3. Pongrácz E (2007) The environmental impacts of packaging
4. Guardian (2017) "Plastic recycling is a myth": what really happens to your rubbish? Environment. The Guardian. https://www.theguardian.com/. https://www.theguardian.com/environment/2019/aug/17/plastic-recycling-myth-what-really-happens-your-rubbish. Accessed 10 Nov 2019
5. The Guardian (2017) Only 14% of plastics are recycled—can tech innovation tackle the rest? Guardian Sustainable Business. The Guardian. Guard. https://www.theguardian.com/sustainable-business/2017/feb/22/plastics-recycling-trash-chemicals-styrofoam-packaging. Accessed 24 Jul 2018
6. Baldacchino L, Cutajar C (2011) The artisan: a sustainable entrepreneur. In: Green jobs from a small state perspective. Case Studies From Malta. pp 20–32
7. Manzini E, M'Rithaa MK (2016) Distributed systems and cosmopolitan localism: an emerging design scenario for resilient societies. Sustain Dev 24:275–280. https://doi.org/10.1002/sd.1628
8. Manzini E (2010) Small, local, open, and connected: design for social innovation and sustainability. J Des Strateg 4:8–11
9. Wegener C (2016) Up-cycling. In creativity—a new vocabulary. Palgrave Macmillan UK, pp 181–188
10. www.plasticexpert.co.uk (2019) Mbongeni Buthelezi: recycled plastic artist. mbongeni-buthelezi-recycled-plastic-artist. https://www.plasticexpert.co.uk/mbongeni-buthelezi-recycled-plastic-artist/. Accessed 2 Sep 2020
11. Unenvironment.org From cartons to carpets: rethinking plastic waste. News Stories. https://www.unenvironment.org/news-and-stories/story/cartons-carpets-rethinking-plastic-waste. Accessed 4 Sep 2020
12. Trashybags.com (2008) Trashy Bags: Eco-friendly recycled plastic bags and gifts, made in Africa. Eco-friendly Recycl. Plast. bags Gift. made Africa. http://trashybags.org/index.htm. Accessed 4 Sep 2020
13. Design W The Wren Design—Recycled cement paper bags, sleeves and accessories. https://www.thewrendesign.com/. Accessed 6 Jan 2021
14. Euromonitor International (2014) Strategic market research, data and analysis. Euromonitor International. https://www.euromonitor.com/. Accessed 19 Sep 2019
15. Bereamail.co.za (2018) The art of plastic. Berea Mail. Art Plastic. https://bereamail.co.za/136192/the-art-of-plastic/. Accessed 3 Sep 2020
16. CocoCollection (2019) Mbongeni Buthelezi, plastic painting art series. Coco Collection. Coco Collect. https://cococollection.com/en/news/post/plastic-painting-with-mbongeni-buthelezi. Accessed 3 Sep 2020
17. Sleek Magazine (2017) The artist critiquing voting systems through performance art. Sleek magazine. https://www.sleek-mag.com/article/serge-attukwei-clottey/. Accessed 4 Sep 2020

Ginn Bonsu Assibey holds a doctorate in design from the Cape Peninsula University of Technology. His research interests include exploring emerging creative and innovative sustainable design approaches at the grassroots level in society, designing sustainability in climate change, and exploring emerging sustainable advertising and promotional strategies. He works mostly in two research paradigms driven by learning through participation: interpretivism and pragmatism.

Alettia Vorster Chisin is a senior lecturer at the Cape Peninsula University of Technology in the Design Department and is currently active in post-graduate supervision in that Department. Her research interests include practice-led explorations based on printmaking techniques and creative writing to inform design research. She has supervised many Masters' students and is currently engaged in doctoral supervision. Her research areas include design for sustainability, socially conscious design, and a designerly approach to environmental issues relating to climate change.

Landscape: A Holistic Approach to Space

19

Ehsan Keramati Niaragh, Morteza Hemmati,
Mohamadreza Forouzandeh, Seyed Amir Mansouri,
and Nima Rezaei

Where is the life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?

T. S. Eliot

Summary

Understanding and awareness of the concept of space as the platform for the manifestation of human life has continually been a controversial topic among scholars of architecture and urban studies. A review of the literature in this field

E. Keramati Niaragh (✉) · M. Hemmati · M. Forouzandeh · S. A. Mansouri
Department of Fine Arts, University of Tehran, Tehran, Iran
e-mail: hemmati.m@ut.ac.ir

M. Forouzandeh
e-mail: frouzandeh@ut.ac.ir

E. Keramati Niaragh · M. Hemmati · M. Forouzandeh · S. A. Mansouri · N. Rezaei
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Tehran, Iran
e-mail: rezaei_nima@tums.ac.ir

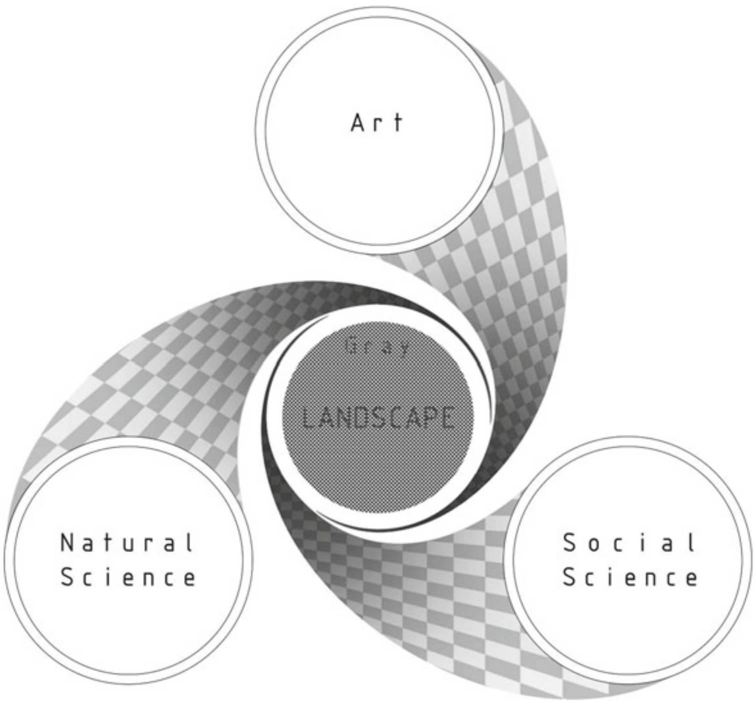
E. Keramati Niaragh · S. A. Mansouri
NAZAR Resaerch Center, No. 23 Dr. Qarib St., Azadi St., Tehran, Iran

N. Rezaei
Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

N. Rezaei
Research Center for Immunodeficiencies, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_19

demonstrates that numerous scholars have attempted to interpret the concept using different knowledge areas, including natural sciences, social sciences, and art. It has led to distinct interpretations of the concept. However, evaluating these findings reveals that each of these approaches, due to its rational base, can partially explain the concept of space, which is not considered a comprehensive insight of the the space as a subjective-objective concept. On the other hand, it seems that landscape, being formed from the interactions of three other approaches, can offer a holistic approach toward the entirety of space, leading to the most comprehensive interpretation. This chapter attempts to evaluate and compare four approaches toward the space concept to determine the most accurate and practical one. It reveals that all three approaches of natural sciences, social sciences, and art lack the conceptual bases required to evaluate the space concept accurately. While, over 500 years, the landscape approach evolved into a concept that can simultaneously interpret the subjective and objective aspects of the space, making it, at least at present, the best-suited approach for the holistic study of this phenomenon.



A landscape approach resulted from conceptual interactions, having various approaches of social sciences, humanities, and arts that make it a multifaceted one, can present a holistic approach that simultaneously considers subjectivity and objectivity of space (subjective-objective).

The code of this chapter is 01001100 01100001 01101110 01100100 01110011 01100011 01100001 01110000 01100101 00001101 00001010.

1 Introduction

Consider the phenomenon of "city" as one of the finest achievements of human civilization. How can the city be described? Is it an arbitrary sequence of the letters comprised in the word? Or perhaps descriptions such as: "*an empty space surrounded by masses of structures,*" *"a societal space for promoting civilization,"* or even *"a sky blotted out by a mass of looming shadows"*? However, which description offers a correct interpretation? Considering that the attempt man has made to describe the phenomena surrounding himself is limited to language, the "point of view" of a phenomenon becomes particularly significant in this regard. However, the evolution of human knowledge through its 2.5 million-year odyssey demonstrates that it has diversified into a multitude, different branches to be able to describe phenomena, rather than bringing forth a uniform and holistic description of them.

Schemata[1] are constantly developing and expanding as human beings continually describe identical phenomena in numerous ways. Modern academicians have reached a consensus regarding the division of human knowledge into a trinity of the natural sciences, the social sciences, and art except for philosophy. It is a part of an endeavor to ontologically and epistemologically examine knowledge's nature. Hence, scholars in each area might investigate the same phenomenon and arrive at different, sometimes contradicting definitions. Each of the descriptions mentioned above of the phenomenon of "city" could represent the predominant approach in each knowledge field. The description *"an empty space surrounded by masses of structures"* could represent the natural sciences relying on visual elements while *"a societal space for promoting civilization"* can represent social sciences. The last description, "*a sky blotted out by a mass of looming shadows,"* represents an artistic approach as suggested by its extravagantly dramatic and metaphoric tone. These examples could demonstrate man's endeavors and experiences regarding space interpretation, aiming to understand his surrounding environment better. A glance at the historical evolution science in its three main fields and the emergence of specialized subfields concerned with the study of space such as ecological and geographical sciences, sociology of space, environmental art, etc., further reveals humanity's quest for understanding and interpreting his surrounding environment.

Moreover, when issues of a more specific space such as urban space are to be studied, areas of study such as "urban planning," "urban engineering," "traffic sciences," "urban sociology" need to be employed as specialized divisions of the main three scientific fields. Despite their superiority in explaining certain aspects of space, each field and their approaches have deficiencies as well, and none of each could present a comprehensive, "holistic" understanding of space. In light of the points elaborated, "Landscape," as an independent discipline with a novel perspective formed through its evolutionary history, aims to introduce the objective

[1] Schemata are packets or maps of data based on an individual's experience regarding their environment which is systemized through continual evolution and revision to optimize data reading and to improve suggested options by a cognitive analyst [1, 2].

and the subjective understanding of space in a uniform and “holistic” approach. This chapter seeks to delve into “landscape” as a “holistic” approach toward the space concept by investigating each main branch of science and its atomistic approaches to space.

2 Natural Sciences

The first area of human knowledge to be discussed is “natural sciences.” However, before proceeding, it will require us to contemplate what type of human knowledge can be considered science? Moreover, to which division of science does natural science point? These are some of the most fundamental questions of the human mind since the beginning of our existence. Following the Modern Period, they have branched off into a fully fledged subdivision of philosophy called “philosophy of science.” Until now, the continuance of this academic discipline demonstrates man’s inability to define the concept of “science” fully; however, many thinkers have strived to draw a distinct line between science and other human knowledge areas since The Golden Age of Greece. The initial great thinker was Aristotle, who divided human knowledge into natural sciences and metaphysics. Modern science as we know it, however, came into being in the seventeenth century, where its foundations occurred on the observation of objective facts. What made the emergence of modern science possible was the immense contributions of great “experimentalists” such as Galileo and Newton, who strived more for experience and observation [3]. Modern science thus focused on observation. Subsequently, science successes throughout the seventeenth, eighteenth, and nineteenth centuries were so awe-inspiring, leading to the foundation of a school of thought known as “logical positivism” by a group of thinkers from the University of Vienna, known as the Vienna Circle, in the 1920s. Logical positivism, itself a by-product of modernism, asserted that the only way a statement could be proven scientific was through direct observation or logical proof. Through the minutely devised standards of methodology introduced in their manifesto, logical positivists not only pioneered the concept of “scientific methodology” [4] but also completely erased the remnants of metaphysical discourses in the realm of science, albeit, temporarily [5]. Consequently, humanities and other study areas founded in abstract concepts rather than concrete facts observable and testable in the physical world were by degrees cast out from the scientific realm.

Despite the numerous remarkable breakthroughs of science from the Early Modern or Renaissance Period to the Modern Period, excessive emphasis on observation and the supremacy of objectivity in the Modern Period led to humanity’s gradual detachment from its living environment. This is due to the natural sciences’ atomistic approach toward phenomena as systems infinitely breakable into smaller parts whose totality can then be examined and determined by studying its comprising parts through observation and experiment [6]. Being one of the two main schools of thought in the objectivist paradigm, this point of view is referred to as "inductivism" in scientific philosophy. The other school, "falsificationism," emphasizes objectivity and testability of observations. However, it did not attribute the generation of identical results from a single phenomenon to the said results being irrefutably true. Falsificationists, thus, called for constant skepticism toward scientific findings on the part of the scientific community. Therefore, natural scientists can only consider each observation as the best temporary option on which they could rely. Accordingly, the more general and falsifiable, yet unrefuted, an assumption, the more scientifically valuable it is. There is, however, a later group of falsificationists who cast doubt on the truthfulness and reliability of observation itself and believe that the scientific community relies too much on observations in experiments and research. Thus, Karl Popper describes the falsificationist notion of the unreliability of observation: *"The empirical basis of objective science has thus nothing "absolute" about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or "given" base, and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being"* [7].

According to the proponents of this branch of falsificationism, observation does not precede understanding, and in scientifically driven observations, human understanding, and precedes observation. Consequently, upon observing identical phenomena, different individuals cannot make the same observation despite the raw data being objective and identical. It lies in the process of understanding that works as a filter or vessel dealing with raw observational data and varies from individual to individual.

In the second half of the twentieth century, the notion, which is still widely popular, observation is the only key to science and research, was challenged and criticized. Post-modern thought prioritized human subjectivity over culture, history, and the environment.

The same historical trend is observable concerning space. Thus, it has been examined as merely another natural world element in an atomistic and divisionary manner. The emergence of new disciplines such as "traffic science," "urban ecology," and "urban planning" can be attributed to the hegemonic dominance of the atomistic point of view toward the phenomenon of the city in the Modern Period. As post-modernist culture came into being and the consequent emergence of new concepts regarding space such as "place," the objective and atomistic perspective of modernism toward space and the division between objective space and the subjective human understanding of it began to fade. As a result, space could be approached from a humanistic perspective based on humanity's understanding of history, culture, and the world. Such could occur free from natural sciences' restraints because their unquestionable reliability [8] and superiority over other kinds of knowledge was challenged. It suggests that relying solely on the natural sciences' empiricism when examining space would reduce mere structural objects. Consequently, as constituent elements of space, landscape, and spatial structures will be regarded as the "totality" of space.

3 Social Sciences

The choice between "social sciences" and "humanities" for this chapter's title proved to be a challenge reminiscent of the dilemma faced by the scholars of this area throughout its evolutionary history[2]. Wilhelm Dilthey (1833–1911), the founder of modern humanities, upon his categorization of the different branches of science, pointed out that social sciences, regardless of their titles, i.e., spiritual, cultural, or social sciences, are distinct from natural sciences. Thus, there exist different methodologies governing these sciences when serving their goals. *"History, political economy, the sciences of law and of the state, the studies of religion, of literature and poetry, of art and music, philosophical worldview"* are revolving around the same ultimate goal, that is, the understanding and the study of a single subject: man [9].

The controversial disagreements between scholars who conservatively stay within the limits of reproducible objective facts and those firmly attached to theoretical syntheses founded upon historical and ethnographical beliefs have been perpetually going on. Social sciences claim to be test-retest reliable like the natural sciences. They can refute or confirm the link between two individual or social phenomena through empiricist experimentation without unfounded explanations. Despite this, social scientists have been progressively concluding that individuals' experience plays a greater role in their behavior and feeling than environmental factors. Moreover, current methods of measurement are not accurate and effective enough to precisely examine the private understandings. It has led to the emergence of a problematic tendency for the quantification of all constructs [10]. As Dilthey points out, the ultimate human sciences subject is the human spirit and mental states, which he refers to as the objective reality of understanding or lived experience [9]. One can assume that a great part of human phenomena is subjective, which led to the development of sequential research methods as relative measurements suitable for qualitative research. Despite this, some scholars tried to discover a link between individuals' biological and psychological states as this proved to be easier to quantify than other methods. The abovementioned narrative of the emergence of the social sciences, or human sciences, demonstrates how natural sciences forced their observation and induction-based methods on the former to be allowed into the realm of science. Despite this, as social sciences developed further, it became evident that there are other qualities and aspects of human existence whose study and understanding would not be possible with conventional scientific methods and would not bear any scientifically valuable results in terms of practicality [10]. As Dilthey points out, the human world is filled with values and meanings while the natural world—of the natural sciences—does not concern itself with such things [9]. In other words, social sciences had to make the sacrifice of limiting their area of study as not all of it was compatible with the

[2] The significant difference between the two fields is related to the function and critique of productions; however, to concentrate on the arts field in this chapter, this section was considered as a social science, which varies from academic one, including sciences related to humanities as well.

scientific method. This shows that social sciences are not alone sufficient to examine all aspects of human phenomena effectively. It has to filter out human and social phenomena. We cannot appropriately measure them through scientific methodology.

In light of the characteristics discussed above, social sciences tend to have a subjective approach toward space and focus on space's human or social components. The expansion of the territory of humanities regarding their qualitative and quantitative methods has led to the development of new combinations in the study of space, which has further enriched the terminology of space and helped shape new insights. Upon perusing the recent literature about space, one encounters various novel combinations such as the study and reading space as language or the role and function of space in fulfilling human needs by providing security, identity, and behavioral norms. Other research areas such as the study of space as personal, social, private, or collective territory also owe their existence to space with concepts borrowed from social sciences. As such concepts found their way into the literature of urban studies as the collective living space for human beings, previously unknown and unexplored aspects and dimensions of urban space were revealed to experts and scholars of the field, ultimately leading to better-informed decisions on planning and executing levels.

4 Art

Artists who create artworks are mere individuals rather than scientists. It seems that the ambiguity, intangibility, indefinability, and the intimation of transcendence found in work of art have impressed upon scientists that they are not adequately equipped to examine art [11]. Moreover, science has not generally shown great interest in art. For philosophers, however, art and properly explaining it has been a subject of great interest and labor. One of the first such attempts was Plato's, who dismissively categorized art as poiesis[3] or mimesis, i.e., imitation: a mere representation of reality. When his "Theory of Forms," an idealistic theory asserting that all physical objects are mere reflections or copies of ideal originals in the world of ideas, is taken into consideration, his dismissive and occasionally hostile treatment of art is only logical as he believed that the most masterfully crafted work of art to be a mere copy of an object in the physical world which itself is an inferior copy or reflection of the ideal [12]. Aristotle was not as harsh in his art treatment and found therapeutic and utilitarian functions to stimulate and purify dangerous sentiments through Catharsis safely. The interpretations as mentioned above of these two great Greek philosophers regarding the essence of art left such a colossal impression on

[3] Aristotle believed art to be a part of the three formal and abstract activities of humanity and their relevant results. Theoria: which is concerned with the theoretical knowledge of man and his understanding of the relationships between things, praxis, which refers to man's activities in order to satisfy his desires, and poiesis, which is an unoriginal and imitative activity concerned with representations of the outside world. Aristotle attributes art to this category [12].

western thought that the duality of "form" and "content" continues to haunt critics and artists alike to this day [13].

Granting all this, when an artist begins the creation process, he or she forgets objective reality and strives to create art, free from the restraints of practicality, resulting from a previously encountered phenomenon as a cohesive whole. In reality and on the contrary, art begins when reality, history, or human experience, cannot be communicated through the employment of linguistic signs, scientific, logical, and philosophical conceptualizations. As Mansouri points out: *"The complexity of understanding work of art is due to its holistic nature,"* which artists achieve with reliance on intuition and revelation [14]. As a result, attempts to understand and interpret this holistic phenomenon have continually dealt with fundamental challenges despite the scholarly community's great breakthroughs in poetry, music, painting, etc. Baumgarten asserts that the reduction of phenomena to their basic substance and components is futile. It might provide insight into each component but will do little to help us understand the whole [15]. Heidegger, too, preferred a holistic approach and warned against the danger of deviation from truth through endeavoring to theorize work of art [16] methodologically.

As discussed earlier, even though art lacked compatibility with scientific standards, due to the influence of art and artists on society, scholars could not be completely dismissive of them with their modernist scientific standards of valuation. Consequently, substantial effort has been made in the global academia to better structuralize and organize theoretical inquiry into the subject; steps toward understanding art in the different areas of the two natural and social branches of science to logically and scientifically regulate art study. Naturally, these attempts, parallel to the rapid expansion of the territory of the social and natural sciences and their novel quantification methods, brought the study of art into a new reductionist stage. This objective point of view endeavored to find meaning and beauty outside the limit of human understanding in the physical external world. Although inductivist and quantitative methods in certain fields of the humanities and art were helpful, this was not true about all of them. As a result, branches with qualitative methods developed in that period and led to serious discussions concerning art's philosophical and psychological aspects. Moreover, natural scientists and researchers focused their studies on the techniques and factors contributing to artistic creation. Furthermore, the progressively significant role artists played in modern society, the functionalist aspect of art became yet another reason for art to be admitted into the world of science as something to be studied for its social, behavioral, neurological, etc. significance [14]. What is remarkable is that the artist continually performs the act of creation through the intuitive employment of their subjective and physical totality of self rather than through rationality and logic, which makes the understanding of art impossible through only natural and social sciences. All contrasts, such as un-learning and learning, forgetting and remembering, uncertainty and certainty, have to be equally considered during the creative field. Hence, during artistic work, the artists have to forget all of the academic processes they have learned so far [17]. The artist has to refer to his existential experiences, rooted in unconscious touch, to create an artwork.

Now to perceive space through the lens of art, phenomenology can be of great help. In an artistic experience, a unique interaction occurs; the artist marks space with their desires and understands the essence of space that engages in a dialectic interaction between their objective and subjective selves and space. Such a process unites objective and subjective structures, merges them, and gives our existential experience cohesion and profound meaning. As Pallasmaa points out: A work of art is thus a series of compact reflections of the experience of existing in the world [17]. Suppose artists when become aware of themselves and their city through this bodily and existential space experience. It leads to an internalization of the urban space in them. This fluid and immediate perception manifest itself in color, light, shadow, musical echoes, etc., as a sentimental representation of space in the artwork.

5 Landscape

The landscape is a relatively new discipline among the abovementioned territories of human knowledge, namely science, humanities, and art. Yet, its interdisciplinary situation distinguishes it from most other disciplines. Its methodological framework is scientific, has goals and subject matter in common with humanities, and its creation occurs as intertwined with art. This relatively distinct character is attributable to its roots and philosophical evolution throughout the last five centuries. Consequently, to better understand the landscape as a holistic approach, it would be beneficial to assess its interactive relationship with social and natural sciences and art throughout its evolutionary history.

Scholars mostly agree on the birth of landscape both as a word and a concept originated from renaissance-related intellectual changes [18–20]. The landscape was first used as a word in the late fifteenth century when the Dutch intended to describe a natural environment and a painting. In this period, landscape emerged with the advent of the classical modern, and within the aesthetic approach of man toward nature, the physical world's discovery, and the distinction between it and the world of phenomena in Europe [21]. As Roger points out, until then, no perception of what we now call a landscape existed, and no trace of such a concept can be found in the literature. In other words, the landscape was a visual invention of the fifteenth century [20]. Parallel to the invention of the perspective technique and modern geometry principles, which helped artistic representation go beyond mere symbolism and subjectivism, the landscape concept attracted more attention. In this period, the term landscape was employed to refer to a sensualistic and subjective artistic concept synonymous with the "aesthetically motivated selection of a natural scene" by the artist previously not common in the art of painting.[4] This mindset

[4] Prior to the Renaissance Period, artistic depictions of a selected natural scene for aesthetic purposes were not common since the art of painting had been generally reserved for mythological, theological, and abstract themes. Roger attributes the works of artists such as Jan Van Eyck and Robert Campin to the emergence of the concept of landscape in this period [20].

toward landscape, amid the eighteenth-century human dominance over nature simultaneous with the systematic and global destruction of landscapes in modern and industrial cities and the emergence of Romanticism in art and literature—and their strong aversion to the predominant modernization and their proponency of an affectionate point of view toward nature—led to the creation of emotive works of art in which not only natural scenes were selected as subjects but were also charged with "sensualistic subjective values."

In the late nineteenth century, amid the decline of the long dominance of positivist ideas and their insistence on the division between the subject (man) and object (the world), significant changes took place in the scientific world. During this period, relativity physics, through its demonstration that any specific observation is relative to the observer's position, rendered the notion of absolute and immutable objectivity obsolete [22]. The principles of classic modernist physics and their proponency of the separation of subject from the object were consequently challenged [21]. At the same time, the employment of scientific methods in different disciplines for the conquest and understanding of the New World changed the concept of landscape. With the emergence of scientific research, Darwin's findings, and the naturalistic explorations of Von Humboldt, the landscape was systematically defined as an independent and perceivable albeit visually oriented phenomenon [23, 24], and its study became an academic discipline. It was first used in this context by Alvin Oppel [25] as a technical term in cultural and human geography concerned with subjective and social symbolic structures of different landscapes. In 1890, the term "cultural landscape" was first used in geography [26] amid the progressively increasing intermingling of social sciences and landscape.

The most fundamental change regarding the concept of landscape in art and philosophy took place in the twentieth century amid the emergence of phenomenology, which revolutionized our understanding of the relationship between subject and object (perceptive and what is perceived). Scholars such as Husserl, contradicting Descartes and Locke's dualism, demonstrated how the content of perceptual experience goes beyond what is visually perceived but rather includes the context of assumptions, memories, connotations, and predictions, which infinitely enrich experience [27]. Simmel asserts that landscape begins to form when a group of natural phenomena on the earth's surface is perceived in a unified manner [28]. Merleau-Ponty, in his phenomenological examination of perception, repeatedly refers to the example of landscape and its perception [21]. This approach views landscape as a type of place[5] characterized by inseparable and intertwined objective and subjective dimensions. Bernard Lassus confirms this indivisible characteristic of the landscape [30]. Berque validates it by defining landscape as a type of place formed due to the interaction between man and environment possessing biological, ontological, and logical attributes [31]. Swaffield additionally points out: *"Landscape not only represents a perceivable phenomenon which can be seen with the*

[5] Place is a subjective-objective concept characterized with not only quantitative and physical attributes, but also non-physical ones such as the emotional response of its audience. Schultz describes it as a "subjective and objective phenomenon" in a holistic sense [29].

eyes but furthermore represents abstract ideas or connotes vivid imaginations of the mind" [32]. Accordingly, as an indivisible synthetic product of the subjective and the objective, the landscape is a phenomenon whose understanding is only possible through a holistic approach.

Conclusively, the most overt characteristic of landscape, as a discipline, is its scientifically regulated holistic approach toward space. On the one hand, landscape, through its adoption of the three major territories of knowledge, namely: natural sciences, humanities, and art, forms an approach that can be practically beneficial, not only theoretically, in the creation and interpretation of space. This interpretation was favorable to architectural and urban scholars. It defined the space as a summation of historical and physical characteristics without fundamental alteration of perspective to classical space. In contrast, holistic view—As mentioned, one of the characteristics of landscape definition—does not distinguish between two objective and subjective aspects in space but considering them two interpretations of a single truth rather than two independent aspects (Fig. 1). Therefore, in the context of landscape interpretation, the city, as space, is not a mere culmination of physical and objective characteristics but also the perceptual understanding of it through its symbols by its inhabitants.[6] In this approach, urban space is not interpreted solely through the measurement of its physical properties but also through examining the collective memories of it shaped by its inhabitants throughout its history.

6 Discussion

In light of the points elaborated earlier, space, as a phenomenon, can be examined through four different approaches. It is worth pointing out that employing each of these approaches will lead to a different interpretation because they have inherently different rational foundations. The fundamental issue is finding the most effective approach to offer a more comprehensive and accurate description of space as a multi-dimensional phenomenon. As pointed out before, the natural science approach is one of the most commonly adopted approaches in studying space and enjoyed extensive popularity among architecture and urban studies scholars in the Modern Period. However, thanks to its inherent logical tendency to separate the objective from the subjective and paying close attention to the latter aspect of different phenomena, natural science has not yet provided a comprehensive image of space as a subjective-objective phenomenon. Such a scientific approach has led to catastrophic human-environment interactions and serious reconsiderations among scholars concerning this approach's fundamental bases. Moreover, these experiences demonstrated that the classic natural sciences laws can no longer be

[6] Symbols themselves are subjective-objective phenomena possessing formal attributes and semantic dimensions in relationship to the minds of their audiences whose understanding of symbols is possible only through holistically approaching them. In order to interpret urban landscapes, it is therefore vital to have a holistic approach as a great many of the constituent elements in the landscape of a city are symbolic ones.

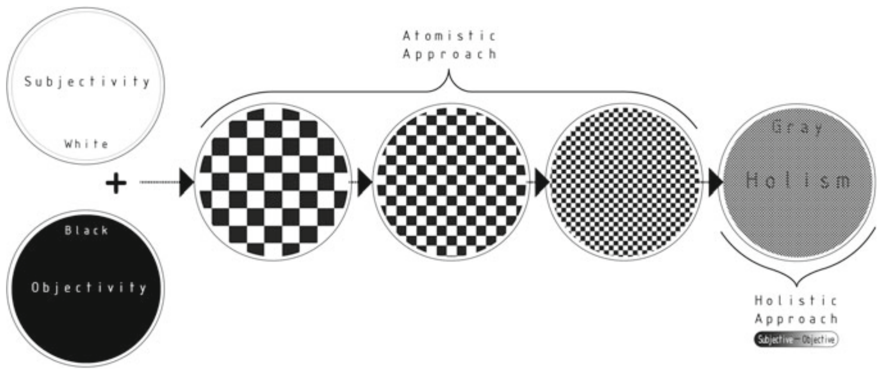


Fig. 1 Atomistic approaches interpret space by integrating its subjective and objective aspects; however, holistic approach does not to distinguish between two aspects, for numerous truths of space synthesize subjectivity and objectivity in new dimension. So, interpretation of this new dimension is solely achieved by holistic approach

held to be absolute and immutable facts and could be subject to revision and change when approached from a holistic critical perspective.

Another popular approach toward the phenomenon of space is that of social sciences, which can consider subjective concepts (or certain subjective concepts) and go beyond the limit of objective reality. However, social sciences can only study concepts confirmable or refutable with scientific methods. Likewise, it excludes subjective issues falling outside this territory. On the other hand, through their simultaneous yet separate examination of phenomena' subjective and objective aspects, social sciences can occasionally offer a divisible sum of the two akin to a separable totality. Contrary to the formerly mentioned approaches, the artistic approach has an inherently holistic tendency toward its treatment of issues and can simultaneously examine the subjective and the objective. Despite that, and because art lacks the scientific methodology required to examine phenomena' quantitative properties, its interpretations of space cannot be as comprehensive and methodic as natural sciences. Finally, the landscape approach as a product of conceptual interactions between the natural and social sciences and art can be regarded as a multifaceted approachable to provide a holistic insight into the objective and subjective aspects of space (Graphical Abstract). Landscape, through a synthesis of these two dimensions, can create a new conceptual dimension referred to as the subjective-objective, enabling us to interpret concepts previously incomprehensive to us, or at best, partly understandable.[7] As a result, one might interpret space as a totality comprised of physical and non-physical dimensions through the landscape

[7] The emphasis of scholars and experts on the uniformity of landscape as a discipline is due to the fact that subjective-objective phenomena are formed in a new context of conceptualization absolutely untranslatable through atomistic approaches. What is represented as an interpretation of a uniform totality by atomistic approaches is not only not a translation of parts of a whole, but also a conceptually different and fundamentally contradictory one.

approach. Such treats space as neither a solely physical phenomenon nor an abstract, subjective issue as there exists a dynamic and concurrent link between the two making them inseparable.

7 Conclusion

Interpreting space has been a perpetual challenge for humanity and has produced numerous mutations of itself under the influence of different approaches throughout history. The study and interpretation of this concept have been subject to the impact of different inquiry fields like natural sciences, social sciences, and art. However, it seems that none of these approaches could provide a comprehensive interpretation of all its aspects and dimensions. As with natural sciences, space reduces to an objective phenomenon whose different dimensions are only measurable and open to exploration through a limited number of factors. Since measurability is an inseparable and fundamental aspect of modern science, the abstract and qualitative aspects of space cannot be adequately examined through natural sciences' employment alone.

Moreover, humanities, committed to studying those abstract and subjective concepts compatible with scientific methodology, seem unable to simultaneously explain the objective and the subjective aspects of space due to its atomistic perspective. Despite possessing the means for a holistic examination of both the subjective and the objective aspects of phenomena, art does not explain all its aspects and properties, namely the quantitative ones, since it lacks methodological means. Nevertheless, the landscape approach, which is a synthesis of the interactions among all three fields mentioned above of study throughout history, has developed certain inherent properties required for holistically approaching and interpreting various and intertwined subjective and objective aspects of space as a multifaceted, multi-dimensional phenomenon.

Core Messages

- Evaluation of various approaches of natural sciences, social sciences, and art reveals that each of these approaches, due to its rational base, can partially explain the concept of space, which is not considered a comprehensive insight of the space as a subjective-objective concept.
- The landscape is an interdisciplinary approach that uses the methodological framework of scientific approach has goals and subject matter in common with humanities, and its creation occurs as intertwined with art.
- The most significant landscape characteristic is its scientifically regulated holistic approach toward space which is attributable to its roots and philosophical evolution throughout the last five centuries.

- The landscape is a multifaceted approachable to provide a holistic insight into the objective and subjective aspects of space which, through a synthesis of these two dimensions, can create a new conceptual dimension referred to as the subjective-objective, enabling us to interpret concepts previously incomprehensive to us.

References

1. Chalmers AF (2013) What is this thing called science? an assessment of the nature and status of science and its methods, 4th edn. University of Queensland Press, pp 1, 6, 113, 131
2. Craig E (2005) The shorter Routledge encyclopedia of philosophy. Routledge, p 943
3. Zimring JC (2019) What science is and how it really works. University of Virginia, Cambridge University Press, p 291
4. Barati N (2004) Holistic and atomistic worldview and their effects on the architecture and urbanism. Baagh-nazar journal, pp 7–24
5. Popper K (2011) The logic of scientific discovery. Taylor and Francis, Routledge, p 92
6. McCain K, Kampourakis K (2020) What is scientific knowledge: an introduction to contemporary epistemology of science. Routledge, New York, pp 5–6
7. Dilthey W (2002) Selected works, the formation of the historical world in the human sciences. Makkreel, R.A, pp 92, 123, 160, 297
8. Kagan J (2009) The three cultures: natural sciences, social sciences, and the humanities in the 21st century. Cambridge University Press, pp 120–131
9. Dissanayake E (2002) What is art for? University of Washington Press, p ix
10. Eldridge R (2003) An introduction to the philosophy of art. Cambridge University Press, p 37
11. Sontag S (2013) Againts interpretation: and other essays. Straus and Giroux, Farrar, pp 10–13
12. Mansouri S (2019) Preface in: research in landscape architecture methods and methodology. Nazar research center of art, architecture and urbanism, Tehran, pp i–iii
13. Cassirer E (1951) The philosophy of the enlightenment. Princeton University Press, pp 338–360
14. Ahmadi B (2008) Truth and beauty. Markaz, Tehran, p 527
15. Pallasmaa J (2009) The thinking hand: existential and embodied wisdom in architecture. Wiley, Chichester, pp 94–130
16. Berque A (2013) Thinking through landscape (AM Feenberg-Dibon, Trans). Routledge, p 3
17. Collot M (2011) La pensée-paysage: philosophie, arts, littérature. Actes Sud, pp 23–25
18. Roger A (1995) Histoire d'une passion théorique ou comment on devient un Raboliot du paysage. In: Roger A (ed) La théorie du paysage en France, pp 424–437
19. Alehashemi A, Mansouri S (2017) Landscape; a shifting concept: The evolution of the concept of landscape from Renaissance. Bagh-e Nazar 14(57):33–44
20. Merleau-Ponty M (2012) Phenomenology of perception. translated from english: Arayan A. J Ketab-e Mah-e Falsafeh 57:120–127
21. Zonneveld IS (1995) Land Ecology. SPB Academic Publishing (2000) Count your blessings? Twenty-five years of landscape ecology. In: Klijn J, Vos W (eds) From landscape ecology to landscape science. WLO/Kluwer Academic Publishers, Wageningen, pp 30-42
22. Farina A (2006). Principles and methods in landscape ecology. Springer, p 10
23. Troll C (1950) Die geographische Landschaft und ihre Erforschung. Studium Generale, 3 Jg., H. 4/5, pp 163–81

24. Jones M (2003) Human geographical landscapes: J.G. Granö's approach to landscapes as scientist and artist. In: Grand O (ed) Origin of landscape science: F.G. Gram and a new pure geography for a new state. The Turku University Foundation, Turku, pp 71–98
25. Carman T (2012) Merleau-Ponty. Translated from English: Olia. M. Tehran: Qoqnous. P 33. In: Rodi F (ed) Vol. Princeton University Press. pp 92, 123, 126, 141, 160, 226, 256, 297
26. Simmel G (2007) The philosophy of landscape. Theory Cult Soc 24(7–8):20–29
27. Mahan A, Mansouri S (2017) The study of Landscape concept with an emphasis on the views of authorities of various disciplines. Bagh- e Nazar 14(47):17–30
28. Berque A (2013) Is the word "landscape" changing There? Landscape Mag 23:25–27
29. Swaffield S (2016) Theory in landscape architecture: a reader (Translated from English by Feizi M). Shahid Rajayi University, Tehran, p 168
30. Arbib MA (1998) Schema theory, the handbook of brain theory and neural networks, pp 830–834
31. Martin J (1984) Toward a cognitive schemata theory of self-instruction. Instr Sci 13:159–180
32. Norberg-Schulz C (2009) Genius Loci: towards a phenomenology of architecture (Translated from English by Shirazi M). Ney Publishers, Tehran, p 59

Ehsan Keramati Niaragh graduated with a BA in Architecture and an MA in Landscape Architecture from the Mohaghegh Ardabili University and the College of Fine Arts (University of Tehran), respectively. After graduation, he had continued designing and research in the Landscape area at Nazar Research Center. He has been an organizing committee member of USERN since its foundation and has participated in Artistic Activities and Consultations (Designing and construction of USERN Prize). He collaborated in urban development projects at the City of Knowledge (University of Tehran and Tehran University of Medical Sciences). His architectural and landscape projects were awarded/ appreciated in national and international competitions. Ehsan also has publications in the landscape area. He has been professionally active in art performances (painting, graphic design, and statue) awarded/appreciated at national and international levels.

Nima Rezaei gained his medical degree (MD) from Tehran University of Medical Sciences (TUMS) in 2002 and subsequently obtained an MSc in Molecular and Genetic Medicine and a PhD in Clinical Immunology and Human Genetics from the University of Sheffield, UK. He also spent a short-term fellowship in Pediatric Clinical Immunology and Bone Marrow Transplantation in the Newcastle General Hospital. Since 2010, Dr. Rezaei has worked at the Department of Immunology and Biology, School of Medicine, TUMS; he is now Full Professor and Vice Dean of International Affairs, School of Medicine, TUMS, and Co-founder and Deputy President of the Research Center for Immunodeficiencies. He is also the founding President of the Universal Scientific Education and Research Network (USERN). He has edited more than 30 international books, has presented more than 500 lectures/posters in congresses/meetings, and has published more than 800 articles in international scientific journals.

Integrated Approaches to Land Management

20

Marta Jaskulak

We must choose to prioritize long-term sustainability over short-term gratification and calculate the true cost borne by societies in the future instead of just the price of actions and policies today. The global community has a responsibility to facilitate this transformation, and it starts by recognizing the environment as a key determinant of human health.

Sania Nishtar

Summary

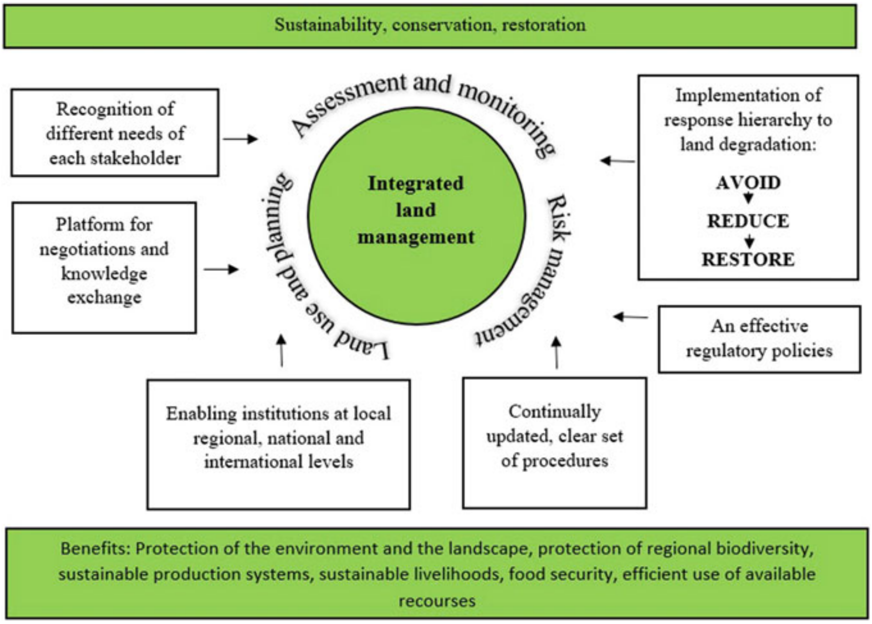
Land degradation (LD) is a process that lowers the quality of land for productive purposes and usually emanates from wind, and water erosion, salinization of soil, logging of water, chemical pollution, or when these factors work together. Land degradation has a severe impact not only on humans but on the entire environment. Integrated approaches to environmental management and planning of land use are currently widely developed and advocated not only by researchers but also by management organizations. The presented chapter focuses primarily on the need for an integrated approach to sustainable land management. It is crucial to implement approaches and techniques developed in the ecological, ecotoxicological, social, behavioral, and economic sciences. Moreover, the proper integration of social and ecological sciences to the

M. Jaskulak (✉)
Laboratoire de Génie Civil Et Géo-Environnement, Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France, ULR4515 - LGCgE59000 Lille, France
e-mail: marta.jaskulak@gumed.edu.pl

M. Jaskulak
Department of Immunobiology and Environment Microbiology, Faculty of Health Sciences with Institute of Maritime and Tropical Medicine, Medical University of Gdansk, 80-210 Gdansk, Poland

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_20

management of ecosystems can broaden our understanding of the long-term dynamics of ecosystems and make the shift from land exploitation to neutrality and finally to the sustainable use of Earth's resources, as well as a shift from inequality, poverty, and hunger to more sustainable development for all citizens of the world.



Integrated approaches to land management

The code of this chapter is 01001100 01100001 01101110 01100100 00001101 00001010.

1 Introduction

Land degradation (LD) is a process that lowers the quality of land for productive purposes and usually emanates from wind, water erosion, salinization of soil, logging of water, chemical pollution, or when these factors work together [1]. Land degradation poses a severe threat to human health and the well-being of the entire environment. Currently, the degradation of land is a continually growing global challenge that converts marginal lands into wastelands and destroys natural ecosystems. The instant triggers for land degradation include deforestation, reduced water resources, destructive land utilization activities, overutilization of synthetic fertilizers, chemicals, pesticides, and industrial and domestic waste disposal [2]. The fundamental factors facilitating LD include rapid population growth, the exploitation that overuses natural resources and accelerated and often ineffectively controlled industrial and urban growth [3].

According to the United Nations (UN), approximately 70% of the 5.2 billion dryland hectares utilized in agriculture worldwide have already suffered some degradation. It affects roughly 250 million individuals globally. Some approximations intimate that the true number of endangered individuals is more than four times higher. For instance, the global arable land area of each individual has experienced as much as 25% shrinkage in solely the final quarter of the twentieth century [4]. Shrinking of land poses severe consequences for food sufficiency and the livelihood of individuals that depend on destroyed lands. The United Nations Environmental Programme (UNEP) approximates that roughly 65 million hectares of forests were globally lost within a span of only five years (from 1990 to 1995). Additionally, the resulting biodiversity loss is severe at genetic, community, and species levels. These estimates need international community attention and an integrated effort to solve these problems [5].

In 2015, the UN adopted the Sustainable Development Goals (SDGs): a total of 17 goals aimed to be the key points for societies to move from land exploitation to neutrality and finally to sustainable use of Earth's resources, as well as to reduce inequality, poverty, and hunger, and offer a proper education and good life for all citizens of the world [6]. Due to the need for food production and access to clean water, proper land management is essential for achieving most of those 17 goals. Moreover, to achieve the SDGs related to health, water, and food production, an increase in land productivity pressure is to be expected [7]. Hence, avoiding further land degradation is a crucial challenge. Thus, achieving a sustainable balance between the needs of the economy, society, climate, and the entire biosphere calls for a holistic and integrated approach [8]. The peripheral regions are the most vulnerable areas in any ecosystem. A lot of land erosion, soil quality degradation, biodiversity loss, and eventually, productivity loss happens in these marginal—but extreme—value lands. It is especially true for drylands, those with arid, semi-arid, and dry sub-humid areas. These areas need sustainable administrative approaches to safeguard, maintain, and reclaim or rehabilitate these delicate environments and their natural resources [3].

2 The Worldwide Implications of Land Degradation

A crucial problem concerning land degradation, particularly in the dry areas, is environmental destruction, which has a close relationship to water quality, which may transform into a very serious and politically challenging issue [9]. The significant loss of land productivity as well as the total benefit loss from the deteriorating ecosystems have significant social impacts and drive the movement of environmental refugees toward the already overpopulated towns and cities. Moreover, the forest land of the Earth is continually declining, the desertification of land is increasing, soils are becoming more and more eroded, and irrigated lands are turning waterlogged and salty. Concurrently, when our planet requires more food for its quickly expanding population, more land is becoming unproductive due to erosion of soil (approx. six to seven ha/year) and the lack of water content or its salinity (one to two million ha/year) [10]. The eroded topsoil gathers in reservoirs, canals, and rivers, which worsens flood destruction, reduces hydropower production, and diminishes the irrigation potential. In particular, poor dryland populations suffer the most severe impacts due to the conversion and the degradation of land. Such problems: deforestation, destroyed rangelands, unproductive soils, salinized soils, and exhausted water sources affect the lives of over 250 million people and pose a threat to another 900 million individuals [11].

The crisis has many complex root causes. Earth and its inhabitants experience a global land administration crisis, with a close link to the water crisis [12]. Hence, the answers to these challenges cannot be individually assessed. As an example, investments in the water sector that aims to alleviate only the flood impact may just worsen the overall situation on that ecosystem, their biodiversity, and the individuals that get their livelihoods from these water systems [13]. Besides, there is a large inventory of corrective actions awaiting to correct the current environmental destruction and rehabilitate the lost biodiversity coming from past industrialization and decades of overexploitation [14]. Previous case studies showed a water problem resulting from inappropriate pricing and allotment guidelines, overdependence on the government for water delivery services, division of water administration between industries and organizations, disregarding health and environmental issues, and regulatory/lawful/organizational collapse in managing both land and water resources [4]. Water resource regulations, strategies, and activities lack sustainability in various countries regardless of the social, environmental, or economic perspective. Local, national, and global environmental challenges have similar principal triggers, mainly the lack of proper economic regulations, policies, and actions, as well as insufficient regulations regarding natural assets and the environment [15]. The problem of handling the complicated network of the global environment and water resources is changing these regulations, minimizes destructive subsidies, and includes environmental needs in the prevailing economic decision making. Past studies reveal that no technological or infrastructural investments will individually work without the need to reform policies, institutions, and laws [7].

It is apparent that land and water resource utilization are closely related, and they must be simultaneously managed with complete biodiversity consideration and with the participation of the communities in each region because each region or even a different part of the region can have different chances, challenges, social organizations, and environmental needs [6]. In the previous years, the “Integrated Water Administration” was developed to offer an answer to those challenges. However, there have been disappointing outcomes because not every sector, consideration for land utilization, or biodiversity have been harmonized with the water sector’s activities. Nevertheless, a crucial lesson from past decades indicates that a complete strategy is required at both the national and regional levels to attain a more strategic set of policy changes, integrated regional management programs, and projects that supports the coordinated response to worldwide agreements that may provide the solution to land and water crisis [16]. The challenge of integrated water, land, and ecosystem administration can only be solved if the proper actions are developed and implemented in each affected region [17].

Thus, there are various duties of the national administration scale. The small scale has different duties, and the local scale also contributes differently by managing responsibilities on the communities, organizations of irrigators, or urban water consumption [5]. This methodical foundation in the management of water and soils (the world bank's water policy calls it a complete foundation) can promote inter-sectoral discussion between different stakeholders to obtain collective advantages for all the regional sectors, considering the biodiversity, instead of triggering sectoral competition [18]. Thus, domestic regulations and activities must simultaneously work with international regulations and relationships between independent nations as globalization continues [19]. Appropriately handling the associated deliberations of water, land utilization, environment, and reducing poverty is a complete foundation that may promote the shift to a new growth model, which will offer constant welfare enhancement and quality of life without spoiling the global ecosystems on which all our economies and lives depend [20].

3 The Need for an Integrated Approach Toward Sustainable Land Management

Sustainability is a prerequisite for the long-term survival of humans. Thus, the subject of sustainable land utilization is becoming more and more vital because of the rising environmental challenges, which include the increased need for all kinds of natural products, climate change, environmental pollution, progressive loss of biodiversity, landscape stability disturbance, economic integration, the security of energy, the supply of water, and disagreements between political–economic, sociocultural, and environmental objectives [21]. There are numerous models and definitions of sustainable land utilization to a worldwide extent, all of which depend on various factors [22]. As an example, the Brundtland report has the most often quoted definition of sustainability: "our single future"—"growth that satisfies the present needs without destroying the ability of the coming generations to satisfy their requirements" [23]. Currently, land sustainability is the basis of the current leading worldwide scheme for international collaboration, which the 2030 Agenda for sustainable growth and its Sustainable Development Goals (SDGs) describes [3]. Numerous definitions emphasized that sustainable growth needs strong socio-economic growth, which observes sustainable land utilization principles and regards the natural and cultural–historical capital and the given area's capability [2]. Overall, sustainable land management should focus particularly on the two newest SDG sustainable land utilization goals. One is objective 15, titled living on the land, and the other is objective 11, called sustainable cities and communities. Since sustainable land utilization promotes stopping or minimizing land destruction and natural threats, it also reduces the loss of biodiversity and promotes the landscape's stability [8]. The administration of sustainable land utilization can be divided into five primary needs as follows;

- the need to ensure and/or improve the region's spatial stability;
- the need to protect nature and considerate natural resource utilization, particularly the soil, water, and the total biodiversity;
- the need to preserve cultural and historical capital and its heritage;
- the need to regenerate human capital and safeguard people's health; and
- the needs of humanization and beautifying the landscape [24].

Sustainable growth emphasizes caring for the landscape by practicing the principles of sustainable living and merging growth with conservation: the conservation of human activities within the capacity of the Earth and growth to allow individuals everywhere to have long and healthy lives [25].

The connection between the land utilization and the land cover has to be addressed firstly in any sustainable management strategy since any kind of land utilization triggers changes on the land cover. Therefore, it is critical to study land utilization and changes in the land cover. Notably, it needs to estimate all the land utilization alteration drivers, the location and relationship of landscape components, causality, the outcomes, and the effects of such alterations [26, 27]. The significant factors that can promote land utilization changes can be: technological, political, economic, natural, and cultural. It is impossible to assess and suggest maximal land utilization based on a single landscape attribute. It thus necessitates examining the link between the various land parameters and that all decision making should use a harmonized technique based on the understanding of landscape as a geo-ecosystem [28]. Harmonized and sustainable land utilization administration approach depend on the previous evaluation of the natural resources and human association facets of the structure of landscape through the right landscape assessment techniques, including a multi-scale analysis and modeling [29]. The impacts of a single land utilization change over a given period determine the sustainability rate of the study area and the relationship between nature and social subsystems. Management of sustainable land utilization must, therefore, strongly rely on the harmonized landscape studies in the three primary facets: environmental, social, and economic [30]. Besides, analyzing the links and ale dependencies between these three facets should have the final objective of defining the management type to monitor socio-economic land utilization progress and preserve its natural, cultural, human, and historical capacity [31].

Expansion in various industries such as agriculture, tourism, production, energy, and mining progressively put pressures on the natural environment and services the ecosystem provides. The steep plummeting of the ecosystem services and the extent of Earth's biodiversity loss is gaining much recognition. Although the rate of biological degradation is slowing, habitat and species loss persists. Industries and conservation firms, including the society at large, appreciate the need to recognize and rectify the loss of ecosystem function and services resulting from inappropriate land management activities [30]. Restoring the ecological ecosystem is thus vital in preserving the environmental, social, and economic values of the societies. Therefore, this will need planning changes, management, control, and a good understanding of reporting land use [32]. Presently, we have an inadequate capacity to determine improvement, preservation, or reduction in our ecological state. The development toward a complete understanding and better monitoring has faced constraints from interaction complexity and inadequate information. Our understanding of ecosystems and the facilities they offer is still limited, coupled with incomplete data for evaluating ecological sustainability and sustainable use. This reduces our capabilities to measure environmental changes and the consequences of modifications in practices of land management [33]. Researchers suggest that if we want to manage natural resources at the regional levels sustainably, we should integrate social and economic information with the learning results adopted from observing the natural resource foundation (Fig. 1). Besides, the knowledge discrepancies regarding the measurement and combining of environmental, social, and economic challenges in the planning process need to be addressed before any action. Moreover, the knowledge gaps in managing the effects on our ecosystems and how to appropriately control, assess, and report on the environment's performance also need more recognition [34].

There should be more encouragement to adopt integrated, cross-sectoral techniques to sustainable resource management and biodiversity conservation within ecologically appropriate planning and administration [35]. Land administrators and policymakers should embrace the harmonized, cross-sectoral approaches inherent in a strategy of ecosystems if they are to contain unwanted or extreme changes in the environment and, eventually, restrict, or even reverse the biodiversity decline. We

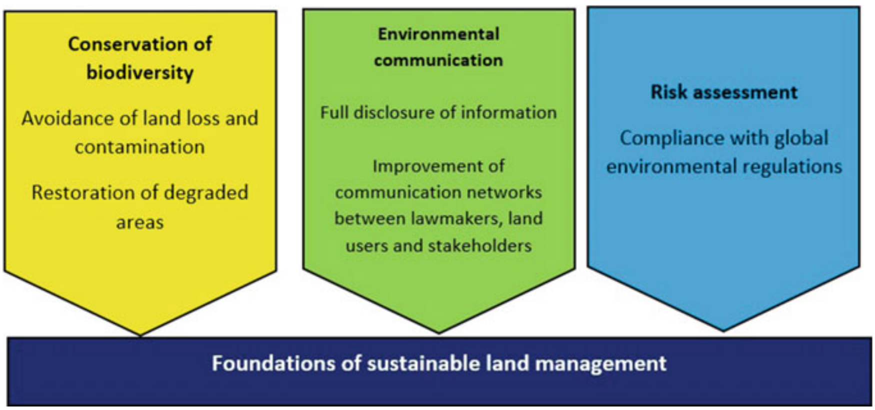


Fig. 1 Basic foundation of sustainable management

can significantly improve the performance and the adeptness of managing our environment by planning and monitoring the utilization of our natural resources while respecting the inherent carrying capacity limits determined for an ecosystem that functions healthily [36]. More development toward a sustainable approach to managing land and restoring ecological functions would be attained by continuously developing and merging these three components that create the ecosystem approach: (a) the bioregional planning and administering of land depending on the units of the ecosystem, (b) the environmental management adaptive systems that tackle the major elements of ecological tenability and results to perpetual enhancement in monitoring the performance of environmental systems, (c) assessing and reporting within integrated bioregional planning and administration plan (Fig. 2) [37].

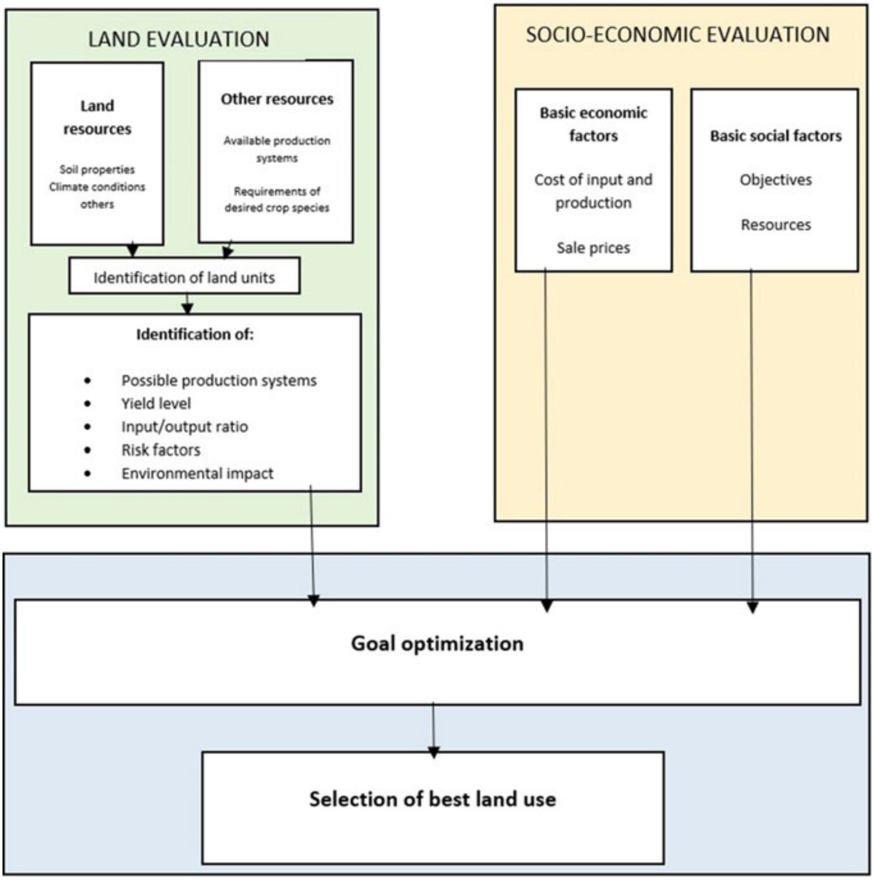


Fig. 2 Basic structure of a decision support system for sustainable land use

4 The Crucial Aspects of an Integrated Approach Toward Sustainable Land Management

Since it is essential to develop integrated approaches to reduce land degradation and the associated societal and economic consequences, it is necessary to encourage building actions and bolster the current institutional abilities for regional, national, and basin-level institutions to handle cross-sectoral factors appropriately [38]. Nevertheless, it is a complicated assignment to define such a harmonized approach, and each region would have a different outcome. These four aspects of the problem must be considered to develop a general foundation for such harmonized approaches [39].

4.1 Technical Aspects

- All inexhaustible natural assets (soil, water, and vegetation) should be considered when creating harmonized administrative plans;
- Ingenious answers should be determined to control degradation of land, majorly through efficient water utilization and production, and conserving soil;
- Possible disagreements and collaboration between lowlands and highlands should be considered, primarily since highlands and mountains serve as lowland water reservoirs; and
- Trans-ecozone resource attributes should be duly considered-particularly water. Trans-ecozone level organization and disagreement settlement are vital in approaches to enhance dry area resource status.

4.2 Human Aspects

- Localized land ownership and tenure approaches are usually vital in resource conservation;
- Effects on local people's livelihoods should be considered when developing and discussing the strategies to manage resources; substitute community livelihoods that may be affected should be created;
- Indigenous activity impacts on natural assets, both destructive and beneficial, can be considered;
- Where possible, there should be explicit consideration of indirect advantages of harmonized administration; and
- The programs should have an inbuilt conflict resolution mechanism during the execution of the administration approaches.

4.3 Economic Aspects

- Social, environmental, and economic losses and gains should be evaluated to guarantee extended tenability or the integrated approach viability;
- Investing capital in creating contemporary infrastructure, including preserving current and traditional activities, should be conducted; and
- The relationship to national economic growth should be explained.

4.4 The Natural Resource Aspect

Integrated programs should prioritize reconstructing marginal ecosystem lands, and If possible, in situ ecosystem biodiversity conservation should be contemplated. These aspects have a close relationship and should have an explicit consideration to create wholly integrated approaches. Various international institutions are presently striving toward creating such strategies; nevertheless, there are few successful such initiatives. Therefore, it is critical to consider the benefits UNU can confer integration attempts for peripheral lands in dry regions.

5 The Ecosystem Approach

The ecosystem approach offers a theoretical model for merging bioregions, management systems for the environment, supervising, evaluating, and reporting across numerous possible ways of land utilization. Figure 3 illustrates the links between the components [18]. An ecosystem model allows merging the major thematic regions (for instance, biological diversity, inland waters, farming, forests, rangelands, and drylands). It has response mechanisms that enable continuous environmental management enhancement [40]. The model produces enhanced institutional collaboration, better data on the environmental foundation, embracing the best practices, and reporting the environment's performance appropriately [16]. Managing biodiversity and creating a result-oriented performance indicator will require frequent observation of environmental factors at scale ranges. This totals to adaptive administration. The key bioindicators of sustainability can be divided in the following way:

- Biophysical
  - Related to land cover
  - Related to land surface
  - Related to soil quality.
- Socio-economic
  - Food sufficiency
  - Increase in human health conditions

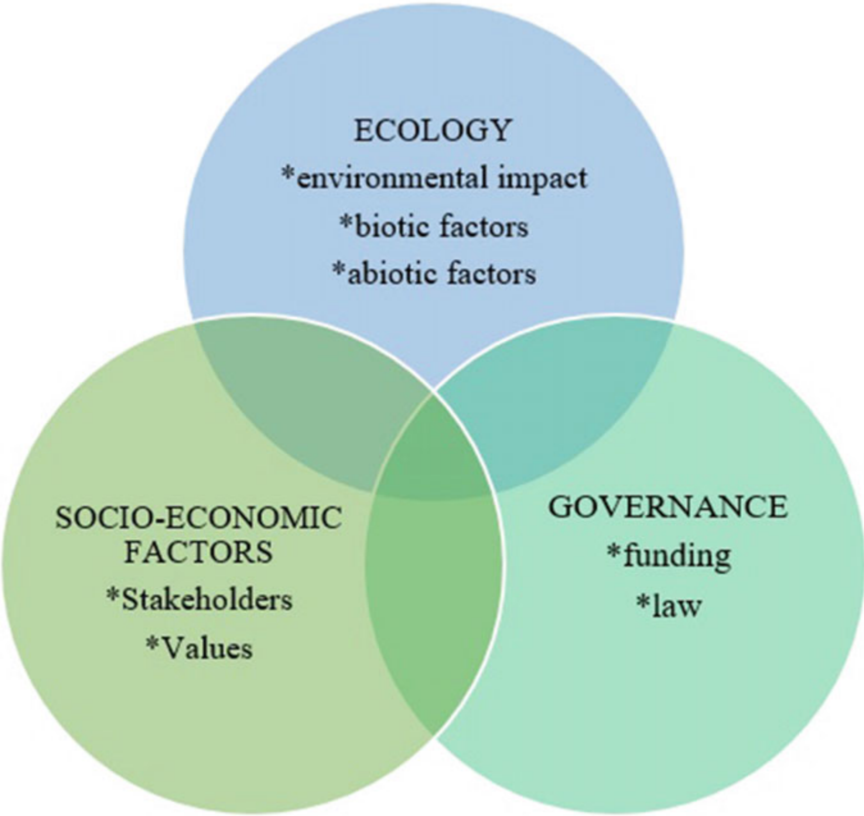


Fig. 3 Three pillars of the ecosystem approach

- Increase or maintenance of food production per capita
- Good relations between different stakeholders and land users
- Maintenance or increase in opportunities for rural labor.

Thus, the ecosystem approach relies on applying the right scientific methods focusing on biological organization levels, including the principal building, processes, responsibilities, and the relationships between organisms and their surroundings [3]. It also appreciates that people, with their diverse cultures, are an essential part of numerous environments. The ecosystem method is crucial in leading activities of the different work programs and in linking the various work strategies [13].

Nevertheless, all biomes, and therefore work programs, have some relationship, and administration attempts will possibly have only a limited success rate if the links are ignored [10]. The ecosystem model needs adaptive administration to handle complicated and changing ecosystem conditions, and the lack of full knowledge or awareness of how they function can severely limit the possible outcomes [12]. Often, the processes of ecosystems are nonlinear. This produces gaps, resulting in surprise and doubt. The administration must be flexible in responding to such uncertainties and preserving the elements of "hands-on-learning" or response from research [14]. Scientific research that seeks to provide an understanding of the broader ecosystem's functioning with regards to its components and their association, and leans toward the need for information by the administration, will make sure that the management's decisions rely on the best available science in the cautionary approach context [19].

5.1 A Bioregional Organization Using Units of Natural Ecosystems (Mapping Ecoregions) for Organization and Administration

Natural resource administrators are increasingly acknowledging that viewing the environment as a specific theme (for instance, soil, vegetation, or climate) does not represent an actual ecological landscape complexity and the various methods they handle the different practices in land administration [25]. They acknowledge the dynamism of ecosystems, continuously reacting to the effects of anthropogenic and natural actions at temporal and spatial range scales. This acknowledgment reveals that it is vital to acknowledge further that ecosystem-based management can be organized to work at various regional levels. Differently said, "ecosystem-based administration" works with the awareness of the guidelines of "bioregional organization." Essentially, ecosystem-based administration needs the synthesis of geomorphology, water value, soil, biota, and vegetation information [41].

The following can describe the principle of ecosystem management: Excellent administration of the ecosystem will preserve the purity of ecosystem functions to prevent escalated injurious ecological or environmental modifications. It will seek to protect and improve the biological diversity and services of the environment, such as the quality of water and supporting the food chain [30]. The ecosystem models give an extensive management basis since the size of the covered administration unit can be altered for various temporal and spatial measures, based on the challenge type and to which scales the ecosystem processes are functioning. Besides, it offers a more reasonable cooperation structure between governments, communities, and personal interests to enable an integrated, interdisciplinary, inclusive, and tenable management approach [36]. The bioregional organization relies on the understanding that ecological systems can be hierarchical, from a large extent (ecoregion) to a smaller extent (communities and smaller constituents). A bioregional planning basis can ensure that different land types and ownership in any area are organized and administered complimentarily to attain long-term conservation of nature, tenable production, and human standards. Put differently. Bioregional organizations harmonize the conservation of biodiversity with the management of natural resources to handle people's expectations and needs [42].

5.2 Management Systems for the Environment

This evaluation framework can also be **implemented** to assess ecologically sustainable agricultural land administration [7]. The Environmental Management System (EMS) scheme offers an adaptive environmental management basis. It is a responsive, process-reliant tool for attaining continuous environmental performance enhancement. It controls the impacts on the environment through the actions of an enterprise using specific procedure sets. These sets can be summed as a "plan-do-check-act" perpetual cycle of enhancement. A vital attribute of EMS is that it can be harmonized in the current bioregional organization. Besides, it can contain any current administration plans at various scales from personal administration units to a "whole landscape." Natural system knowledge is vital in embracive administration and continuous enhancement of environmental functions [26]. Adaptive administration, coupled with the above described bioregional organizational approach, offers an efficient model for detecting and managing natural or anthropogenic alterations in the environment's values or in time or space evaluation. When combined, these two models offer the planner and the administrator a more exceptional ability to enhance the health and the production potential of the whole ecosystem [38]. For preserving the services, it offers the community over short, medium, and long durations. An EMS will adopt various information and standards, regulations, and laws to provide the adaptive process management basis that boosts the continuous environmental performance function and attains sustainable environmental management outcomes [4].

5.3 Monitoring, Assessment, and Reporting

The fundamental significance of an EMS as a flexible and reliable management tool for our natural resources is based on whether monitoring and assessment of its environmental functionality offer a sufficient level of information for the planned activities to attain all envisaged environmental results [2]. Hence, developing performance indexes, as well as the criteria for natural resource management monitoring and sustainable agricultural production, is essential in executing EMS plans. The large-scale biophysical resource inventory requirement of planning ecoregional processes is also mandatory in establishing the monitoring baseline, assessing, and reporting [6]. The obtained data, gathered from the process of monitoring, and defined against a proper control, can give us primary information to support the management decisions to control the degree or direction of change in the state or working of an ecosystem component or activities. It is worth noting that organizing, administration, and policy needs, including ecological requirements, will always affect data collection. Thus, the functional indexes for monitoring results should provide relevant information at temporal and spatial interest levels to public decision-makers and resource administrators [43]. The right organizing and administration authorities should thus define organizational goals and needed results, coupled with steps of attaining these desired results, before definition and then gathering performance indicator data for preserving standards and resources [28]. This need is inherent in the organization and administration tools that establish an EMS. Nationally accepted protocols must be applied for monitoring where the gathered information at the local and regional levels is to be assembled to a national standard for reporting against the indexes [11].

6 Conclusion

Decisions of managing natural resources determine the future well-being and the potential of our ecosystems to produce in several ways, and eventually, their capacity to offer products and services [42]. Although it is the responsibility of private and public policymakers and land administrators to preserve and guarantee the future standard and of these products from the ecosystem, currently, not every policymaker adequately considers biodiversity and products and services from the ecosystem on which our life support mechanism relies [40]. Our natural resource information base has critical disparity, and we have extreme limitations in our knowledge store to manage the natural resources. These pose considerable obstacles to the tangible execution of natural resource administration guidelines and administrative actions. Productive administration needs enough observation to recognize the causal relationships between the alteration factors and reaction at an exceptional measure and then create the right work to modify the existing activities to alleviate any extreme environmental progress [24].

Nevertheless, the quality or quantity of monitoring data does not usually suffice to offer statistically valid interpretations. We have limited capacity to develop, collect and interpret indicator scales, observe changes and progress in the levels, and evaluate their extent against a predefined standard for numerous management challenges. Governments and the community increasingly need information on the progress and natural resource conditions over the whole landscape [12]. If we want to succeed in satisfying this need, we must create and embrace a series of quantifiable results-oriented functional indexes for natural resource administration that can be observed across every land utilization within a concessional bioregional organization and the above proposed flexible administrative approach.

Core Messages

- Decisions of managing natural resources determine our future well-being and the potential of ecosystems to produce goods and services.
- Land administrators are increasingly acknowledging that viewing the environment as a specific theme (for instance, soil, vegetation, or climate)

does not represent an actual ecological landscape complexity and how they handle the different practices in land administration.

- To strike a balance between the needs of the economy, society, climate, and the entire biosphere calls for a holistic and integrated approach.

References

1. Grobelak A, Jaskulak M (2019) Sludge multifunctions in a phytobiome—forest and plantation application including microbial aspects. Industr Municipal Sludge 323–336. https://doi.org/10.1016/b978-0-12-815907-1.00014-3
2. Pacheco FAL, Sanches Fernandes LF, Valle Junior RF, Valera CA, Pissarra TCT (2018) Land degradation: multiple environmental consequences and routes to neutrality. Curr Opin Environ Sci Health 5:79–86. https://doi.org/10.1016/j.coesh.2018.07.002
3. Kust G, Andreeva O, Cowie A (2017) Land degradation neutrality: concept development, practical applications and assessment. J Environ Manage 195:16–24. https://doi.org/10.1016/j.jenvman.2016.10.043
4. Arvor D, Tritsch I, Barcellos C, Jégou N, Dubreuil V (2017) Land use sustainability on the South-Eastern Amazon agricultural frontier: recent progress and the challenges ahead. Appl Geogr 80:86–97. https://doi.org/10.1016/j.apgeog.2017.02.003
5. Yigitcanlar T, Dur F (2010) Developing a sustainability assessment model: the sustainable infrastructure, land-use, environment and transport model. Sustainability 2:321–340
6. Viallon F-X, Schweizer R, Varone F (2019) When the regime goes local: local regulatory arrangements and land use sustainability. Environ Sci Policy 96:77–84. https://doi.org/10.1016/j.envsci.2019.02.010
7. Heck V, Hoff H, Wirsenius S, Meyer C, Kreft H (2018) Land use options for staying within the planetary boundaries—synergies and trade-offs between global and local sustainability goals. Glob Environ Change 49:73–84. https://doi.org/10.1016/j.gloenvcha.2018.02.004
8. D'Amato D, Korhonen J, Toppinen A (2019) Circular, green, and bio economy: how do companies in land-use intensive sectors align with sustainability concepts? Ecol Econ 158:116–133. https://doi.org/10.1016/j.ecolecon.2018.12.026
9. Jaskulak M, Grobelak A, Vandenbulcke F (2020) Modelling assisted phytoremediation of soils contaminated with heavy metals—main opportunities, limitations, decision making and future prospects. Chemosphere 126196. https://doi.org/10.1016/j.chemosphere.2020.126196
10. Barbier EB, Hochard JP (2018) Land degradation and poverty. Nat Sustain 1(11):623–631. https://doi.org/10.1038/s41893-018-0155-4
11. Symeonakis E, Karathanasis N, Koukoulas S, Panagopoulos G (2014) Monitoring sensitivity to land degradation and desertification with the environmentally sensitive area index: the case of Lesvos Island. Land Degrad Dev 27(6):1562–1573. https://doi.org/10.1002/ldr.2285
12. Liebig MA, Herrick JE, Archer DW, Dobrowolski J, Duiker SW, Franzluebbers AJ, Strickland TC (2017) Aligning land use with land potential: the role of integrated agriculture. Agric Environ Lett 2(1):170007. https://doi.org/10.2134/ael2017.03.0007
13. Webb NP, Marshall NA, Stringer LC, Reed MS, Chappell A, Herrick JE (2017) Land degradation and climate change: building climate resilience in agriculture. Front Ecol Environ 15(8):450–459. https://doi.org/10.1002/fee.1530
14. Cowie AL, Orr BJ, Castillo Sanchez VM, Chasek P, Crossman ND, Erlewein A, Welton S (2018) Land in balance: the scientific conceptual framework for land degradation neutrality. Environ Sci Policy 79:25–35. https://doi.org/10.1016/j.envsci.2017.10.011

15. Galford GL, Soares-Filho B, Cerri CEP (2013) Prospects for land-use sustainability on the agricultural frontier of the Brazilian Amazon. Philos Trans R Soc B Biol Sci 368 (1619):20120171–20120171. https://doi.org/10.1098/rstb.2012.0171
16. Hermanns T, Helming K, König HJ, Schmidt K, Li Q, Faust H (2017) Sustainability impact assessment of peatland-use scenarios: confronting land use supply with demand. Ecosyst Serv 26:365–376. https://doi.org/10.1016/j.ecoser.2017.02.002
17. Seghezzo L, Venencia C, Buliubasich EC, Iribarnegaray MA, Volante JN (2016) Participatory, multi-criteria evaluation methods as a means to increase the legitimacy and sustainability of land use planning processes. The Case of the Chaco Region in Salta, Argentina. Environ Manage 59(2):307–324. https://doi.org/10.1007/s00267-016-0779-y
18. Reisi M, Aye L, Rajabifard A, Ngo T (2016) Land-use planning: Implications for transport sustainability. Land Use Policy 50:252–261. https://doi.org/10.1016/j.landusepol.2015.09.018
19. Holman IP, Brown C, Janes V, Sandars D (2017) Can we be certain about future land use change in Europe? A multi-scenario, integrated-assessment analysis. Agric Syst 151:126–135. https://doi.org/10.1016/j.agsy.2016.12.001
20. Hassan MM, Nazem MNI (2015) Examination of land use/land cover changes, urban growth dynamics, and environmental sustainability in Chittagong city Bangladesh. Environ Dev Sustain 18(3):697–716. https://doi.org/10.1007/s10668-015-9672-8
21. Keesstra S, Mol G, de Leeuw J, Okx J, Molenaar C, de Cleen M, Visser S (2018) Soil-related sustainable development goals: four concepts to make land degradation neutrality and restoration work. Land 7(4):133. https://doi.org/10.3390/land7040133
22. Jaskulak M, Grobelak A, Vandenbulcke F (2020) Effects of sewage sludge supplementation on heavy metal accumulation and the expression of ABC transporters in Sinapis alba L. during assisted phytoremediation of contaminated sites. Ecotoxicol Environ Saf 197:110606. https://doi.org/10.1016/j.ecoenv.2020.110606
23. Brundtland G (1987) Report of the world commission on environment and development: our common future
24. Thorsoe MH, Noe EB, Lamandé M, Frelih-Larsen A, Kjeldsen C, Zandersen M, Schjonning P (2019) Sustainable soil management—farmers' perspectives on subsoil compaction and the opportunities and barriers for intervention. Land Use Policy 86:427–437. https://doi.org/10.1016/j.landusepol.2019.05.017
25. Mills J, Reed M, Skaalsveen K, Ingram J (2019) The use of Twitter for knowledge exchange on sustainable soil management. Soil Use Manag. https://doi.org/10.1111/sum.12485
26. Fry P, Thieme S (2019) A social learning video method: Identifying and sharing successful transformation knowledge for sustainable soil management in Switzerland. Soil Use Manag 35(1):185–194. https://doi.org/10.1111/sum.12505
27. Techen A-K, Helming K (2017) Pressures on soil functions from soil management in Germany. A foresight review. Agron Sustain Dev 37(6). https://doi.org/10.1007/s13593-017-0473-3
28. Ahmad Z, Imran M, Qadeer S, Hussain S, Kausar R, Dawson L, Khalid A (2018) Biosurfactants for sustainable soil management. AdvAgron 81–130. https://doi.org/10.1016/bs.agron.2018.02.002
29. Powlson DS, Gregory PJ, Whalley WR, Quinton JN, Hopkins DW, Whitmore AP, Goulding KWT (2011) Soil management in relation to sustainable agriculture and ecosystem services. Food Policy 36:72–87. https://doi.org/10.1016/j.foodpol.2010.11.025
30. Ingram J (2008) Are farmers in England equipped to meet the knowledge challenge of sustainable soil management? An analysis of farmer and advisor views. J Environ Manage 86 (1):214–228. https://doi.org/10.1016/j.jenvman.2006.12.036
31. Brevik EC, Pereira P, Muñoz-Rojas M, Miller BA, Cerdà A, Parras-Alcántara L, Lozano-García B (2017) Historical perspectives on soil mapping and process modeling for sustainable land use management. In: Soil mapping and process modeling for sustainable land use management, pp 3–28. https://doi.org/10.1016/b978-0-12-805200-6.00001-3

32. Hansjürgens B, Lienkamp A, Möckel S (2018) Justifying soil protection and sustainable soil management: creation-ethical legal and economic considerations. Sustainability 10(10):3807. https://doi.org/10.3390/su10103807
33. Seto KC, Reenberg A, Boone CG, Fragkias M, Haase D, Langanke T, Simon D (2012) Urban land teleconnections and sustainability. Proc Natl Acad Sci 109(20):7687–7692. https://doi.org/10.1073/pnas.1117622109
34. Worrall R, Neil D, Brereton D, Mulligan D (2009) Towards a sustainability criteria and indicators framework for legacy mine land. J Clean Prod 17(16):1426–1434. https://doi.org/10.1016/j.jclepro.2009.04.013
35. Verstegen JA, van der Hilst F, Woltjer G, Karssenberg D, de Jong SM, Faaij APC (2015) What can and can't we say about indirect land-use change in Brazil using an integrated economic - land-use change model? GCB Bioenergy 8(3):561–578. https://doi.org/10.1111/gcbb.12270
36. Li Y, Wu W, Liu Y (2018) Land consolidation for rural sustainability in China: Practical reflections and policy implications. Land Use Policy 74:137–141. https://doi.org/10.1016/j.landusepol.2017.07.003
37. Hou D, Ding Z, Li G, Wu L, Hu P, Guo G, Wang X (2017) A Sustainability Assessment Framework for Agricultural Land Remediation in China. Land Degrad Dev 29(4):1005–1018. https://doi.org/10.1002/ldr.2748
38. Song Y, Hou D, Zhang J, O'Connor D, Li G, Gu Q, Liu P (2018) Environmental and socio-economic sustainability appraisal of contaminated land remediation strategies: A case study at a mega-site in China. Sci Total Environ 610–611:391–401. https://doi.org/10.1016/j.scitotenv.2017.08.016
39. Oberlack C, Tejada L, Messerli P, Rist S, Giger M (2016) Sustainable livelihoods in the global land rush? Archetypes of livelihood vulnerability and sustainability potentials. Glob Environ Change 41:153–171. https://doi.org/10.1016/j.gloenvcha.2016.10.001
40. Qi X, Fu Y, Wang RY, Ng CN, Dang H, He Y (2018) Improving the sustainability of agricultural land use: an integrated framework for the conflict between food security and environmental deterioration. Appl Geogr 90:214–223. https://doi.org/10.1016/j.apgeog.2017.12.009
41. Bryan BA, Nolan M, McKellar L, Connor JD, Newth D, Harwood T, Hatfield-Dodds S (2016) Land-use and sustainability under intersecting global change and domestic policy scenarios: trajectories for Australia to 2050. Glob Environ Change 38:130–152. https://doi.org/10.1016/j.gloenvcha.2016.03.002
42. Chen K, Long H, Liao L, Tu S, Li T (2020) Land use transitions and urban-rural integrated development: theoretical framework and China's evidence. Land Use Policy 92:104465. https://doi.org/10.1016/j.landusepol.2020.104465
43. Jaskulak M, Grobelak A (2019) Chapter 6—Cadmium Phytotoxicity—Biomarkers. Cadmium Tolerance in Plants, Academic Press, pp 177–191

Marta Jaskulak is a researcher at Department of Immunobiology and Environment Microbiology, Faculty of Health Sciences with Institute of Maritime and Tropical Medicine, Medical University of Gdansk, 80-210 Gdansk, Poland, and the Laboratoire de Génie Civil et géo-Environnement, University of Lille, France. Her main research interest includes the soil contamination by heavy metals and its influence on plants and invertebrates. She implements molecular biology techniques, including next-generation sequencing, into environmental research to precisely assess the influence of metal contamination on the physiology of organisms. Moreover, other areas of her research include waste management, specifically sewage sludge management, and bioremediation and phytoremediation of post-industrial land.

Archaeology as an Interdisciplinary Science at the Cross-Roads of Physical, Chemical, Biological, and Social Sciences: New Perspectives and Research

21

Derya Yılmaz

When Willard F. Libby first discovered radiocarbon dating in 1947, archaeologists, and especially Egyptologists, ignored it. They questioned its reliability, as it did not coincide with the "known" historical dates of the artifacts being tested. David Wilson, author of The New Archaeology, wrote, "Some archaeologists refused to accept radiocarbon dating. The attitude of the majority, probably, in the early days of the new technique was summed up by Professor Jo Brew, Director of the Peabody Museum at Harvard. 'If a C14 date supports our theories, we put it in the main text. If it does not entirely contradict them, we put it in a footnote. And if it is completely out-of-date we just drop it.

Christopher Dunn

Summary

This chapter will focus on new research in archaeology, interdisciplinary science that uses various physical, chemical, biological, and social sciences. Such studies, also known as archaeological sciences, are generally referred to as archaeometry. Information kept hidden by an artifact from the past can be learned with the help of archaeological sciences. Archaeology is, as an interdisciplinary field of science, fundamentally considered to be a social

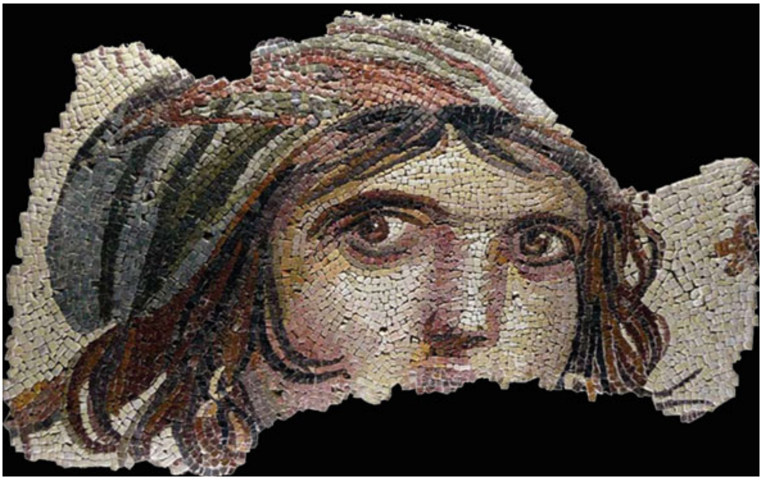
D. Yılmaz (✉)
Faculty of Languages, History and Geography, Ankara University, Ankara, Turkey
e-mail: d.yilmaz@ankara.edu.tr

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Ankara, Turkey

Archaeology Department, Protohistory and Near Eastern Archaeology, Sıhhiye-Ankara 06100, Turkey

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_21

science. However, it is an interdisciplinary science because various basic hard sciences such as physics, chemistry, biology, statistics, maths, and social sciences such as economics and sociology occur in archaeological research. Firstly, the relationship between natural sciences and archaeology and then the relationship between social sciences and archaeology are addressed in the light of new studies. Lastly, the approach of society and archaeology is considered. The number of joint studies and analyses of the natural sciences is increasing every day. For 60 years, the archaeological sciences and archaeology have produced important results with integrated studies in human history. Archaeology uses current methods that are constantly renewed thanks to archaeological science. This allows the integration of any other science in an archaeological study. For instance, computer science's frequent application in archaeological science in recent years is a crucial development. New methods and discoveries in the integrated sciences make the future of archaeology exciting. Thanks to the scientific cooperation in question and with the help of analyses, the distant past will be closer to us and more comprehensible.



"Gypsy Girl" mosaic Gaziantep Zeugma Excavation (Adapted with permission from the archives of Zeugma Excavation).

The code of this chapter is 01000001 01110010 01100011 01101000 01100001 01100101 01101111 01101100 01101111 01100111 01111001 00001101 00001010.

1 Introduction

Archaeology is an interdisciplinary science since the information used in research methods in excavations and obtained in excavations and research is related to other sciences. Archaeology is a word derived from the words "*arkheos*" (ancient) and "*logos*" (path, method, science) and means "*Science of the Past.*" It provides information about the various disciplines of science and also scientific discoveries. Such leads to the development of civilization's history, humanity, and specific ancient cultures while providing insights into human creativity elements. It also includes information about the development of art, trends in art, human relationships with nature, etc.

Archaeology and archaeometry have become inseparable today. The archaeology sciences can help manage it to reach the maximum level of knowledge from an archaeological excavation project. While this precisely answers many questions, it also provides a new perspective illuminating the past's unknowns. Archaeology takes place between experimental sciences and observational sciences. Is archaeology a social science? In its shortest known definition, archaeology studying humans, society, and cultures in the past is also a social science. In this context, it uses various methods and techniques of the social sciences. Today's archaeological projects no longer include only archaeologists and photographers. Archaeological sciences experts participate in archaeological site work and research their subjects personally. This has led to an increase in archaeology's importance in recent years. Accordingly, archaeological site work becomes combined with the fundamental issues of archaeological sciences. Many archaeologists feel the need to analyze archaeological sciences such as geology, chemistry, physics, or biology after unearthing and interpreting an artifact. Thus, an archaeologist can better trace the past than ever by evaluating archaeological science results with site results. The archaeological sciences' methods and techniques are being used together, ranking archaeology first among the interdisciplinary sciences. Indeed, techniques that the archaeological sciences use consist of basic techniques and methods. It primarily borrows the background on basic principles and phenomena of physics of the natural sciences, chemistry, geology, geophysics, astronomy, maths, and information science (Fig. 1) [1].

2 Natural Sciences and Archaeology

For an archaeologist, archaeometry studies are priceless in terms of showing the unseen. For instance, the metal composition of a metal dagger can be identified by the analysis. If there is fabric preserved on the metal, archaeometry studies can be performed for this fabric, and it can provide a better understanding. Analyses applicable to the fabric, for example, are Schweitzer's reagent for cellulose fiber to determine the fiber type and whether it is linen or cotton, Loew reagent to see whether it is natural silk, XRD analysis for mineralized remains, and examination by scanning electron microscope (SEM) to understand their fiber structure in detail [2]. The examination of various remains obtained by sifting soil at the base of a house will provide more information about the house. Archaeological excavations oblige many disciplines to work together. In this context, it has a renewed cycle of methods and theories continuously [3]. While the archaeometric examination of

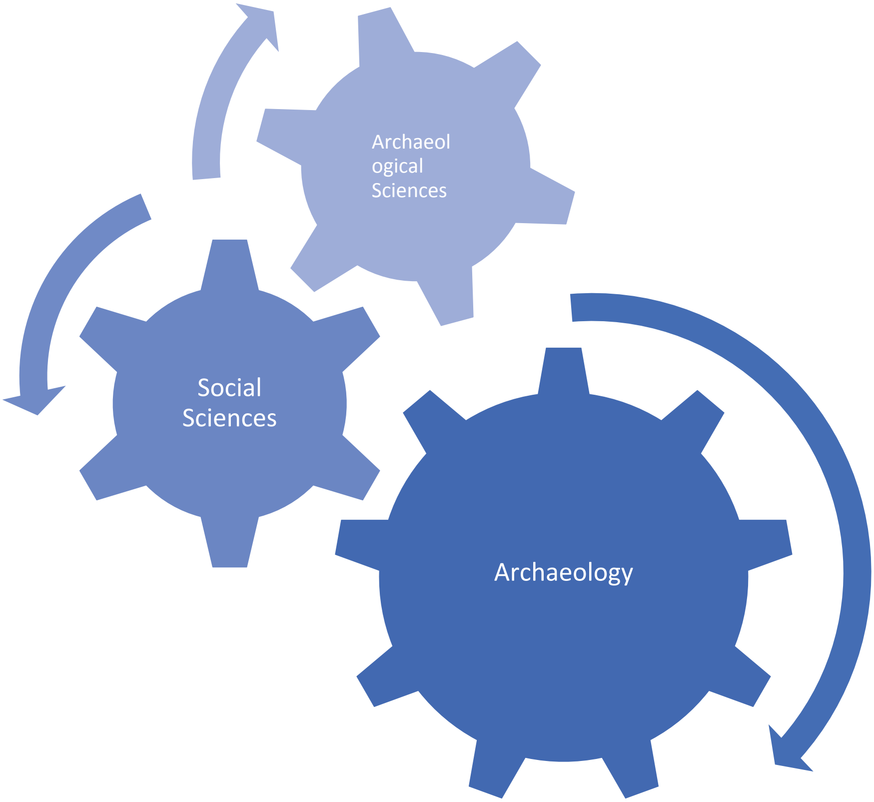


Fig. 1 Archaeometry and other scientific fields. A simple schema shows the relations between archaeology, archaeological sciences, and social sciences

certain artifacts starts directly in the laboratory, for instance, some remains such as skeletons unearthed by anthropologists during excavation, anthropological methods and theories should also be used at the site. Some information obtained from archaeometric studies has illuminating results for today. For example, the glaze analysis concluded that Roman glazed drinking vessels contained lead, and Roman soldiers might have died due to poisoning. Therefore, lead glazing in ceramic technology was prohibited [3].

Archaeometry can apply physics, chemistry, geology, and engineering sciences to analyze various archaeological materials. The term archaeometry comprises the Greek words *archaeos* (ancient) and *metron* (measurement). With the discovery of radiocarbon in the 1950s, modern archaeometry came to exist. Dating an archaeological material using the C 14 method led to a breakthrough in that period's archaeology. It contributed to the development of chronology across the globe. Modern archaeometry was first used as a science and a term in the late 1950s by C.F.C. Hawkes of Oxford University [1, 4]. Since then, archaeometry has always offered new methods and analyses and is continually updating itself. Today, it is also referred to as the archaeological sciences.

The use of natural sciences in archaeology, the determination of the amount of carbon or carbon-14 (14C) in organic materials (wood coal, bones, shells, decayed wood, plant material, etc.), were invented by American W. Libby in the mid-1950s at the Institute of Nuclear Science in Chicago. He was awarded the Nobel Prize in chemistry in 1960 [1]. However, some analyses performed with methods of specific positive sciences date back to the 1930s. Some archaeological studies carried out in the Near East offered pioneering archaeometric analyses. For instance, the Alishar Höyük excavation in Yozgat, Turkey, was one of the most important excavations of the University of Chicago in those years. However, in the 1930s, archaeometric analyses of human and animal bones, seashells, obsidian, wood, ceramics, grains, the fabric remains, and metals were carried out. The publication of the Americans' Alishar Höyük excavation reports supported with archaeometric analyses was a first for Anatolian archaeology [5]. The publication in question is also one of the pioneering studies, including archaeometric analyses in world archaeology.

The use of developments in the positive sciences by their adaptation to archaeology leads to new studies. The network of relations between archaeology and archaeometry varies day by day, and the number of disciplines in this field is increasing. An interaction network between the archaeological sciences occurs when analyzing various materials from an archaeological area based on humans, society, and the environment [1].

Archaeometry offers the opportunity to evaluate archaeological materials for various purposes. There are nine different groups most commonly used. Among them, dating methods with physical and chemical techniques determine the raw material of archaeological material with chemical or physical processes, examine remains under the ground with geophysical methods, and establish a connection between celestial bodies and monuments in terms of astronomy. Moreover, conservation sciences are essential in preserving and restoring monuments with physical and chemical methods. It is valuable determining past environments in terms of fauna, flora, and climate with archaeological science methods in environmental archaeology. Digital archaeology is a new archaeometry area that transfers the past information of an object digitally to the future by allowing this object to be completed thanks to the computer sciences using cyber-archaeology. Digital mapping is also integrated into the evaluation of ancient settlements with all types of mapping using geomatic engineering techniques. These are based on various sub-branches such as lithic analysis, palynology, osteoarchaeology, zooarchaeology, palaeoethnobotany, archaeometallurgy, archeogeophysics, chemistry, bioarchaeology, archaeobotany, meta-archaeology (philosophy of archaeology), conservation and restoration sciences, archaeometallurgy, pathology, ancient DNA, and geomatics and artifact studies (Table 1) [1]. A new method emerging in the sciences mentioned above, archaeometry, is updated continuously by enabling new evaluations. In this sense, archaeometry methods have a continually expanding range. This is a real reflection; the more science progresses, it illuminates the past to the same degree. For instance, archaeological samples' examination has increased in parallel with technological improvements in DNA archaeology within the last 12 years. Mobile DNA labs are on the agenda. In this way, it will be possible to examine samples at the site during excavation or research. According to the results of current studies, ancient DNA studies have clearly shown the determinations

Table 1 Archaeology and integrated science of physical, chemical, biological, and social sciences

Archaeology	
Archaeological sciences	Social sciences
Chemistry (Diffusion, reactions, melting, affinities, archaeometallurgy, obsidian source determination, mineral ore source determination, clay source determination in ceramics and XRF (X-Ray Fluorescence), NAA (Neutron Activation Analysis), ICP-MS (Inductive Coupled Plasma-Mass Spectrometry) FTIR (Fourier Transform-Infrared Spectroscopy) and Raman Spectroscopy)	*Philosophy* (Meta-archaeology)
Physics (Radioactivity, C14 Dating, dendrachronology, luminescence methods, electricity and magnetism, atomic theory, electromagnetic ration)	*Museology* (Exhibition, education)
Biology (aDNA, isotopic analysis, paleogenomics, osteoarchaeology, bioarchaeology, zooarchaeology, palaeoethnobotany, ancient diet and botany)	*Geography* (Human geography, ancient climate, ancient landscape, paleogeography, GIS)
Geology (Geomorphology, geoarchaeology, pedimentology, petrology, paleopathology, archaeoenvironment)	*Sociology* (Social organization, society, social relation, social systems)
Geophysics (Paleoclimate, geomagnetic field, atmosphere, ground penetrating radar (GPR), electrical resistance prospection and magnetic prospection-magnetic susceptibility)	*Economics* (Ancient economy, commerce network, trade)
Mathematical Sciences (Algorithms, statistics, geometry, mathematical modelling)	*Ethnography* (Ethnoarchaeology)
Astronomy (Archaeoastronomy, Solar system, celestial mechanics, sun, moon, symbol)	*Anthropology* (Paleoanthropology, social anthropology, phycical anthropology)
Cyber-archaeology/information science (Information and communication Technology (ICT), software engineering, computer sciences, photography digital archive, virtual Reality, 3D modelling)	*Philology* (Ancient linguistics)
Conservation sciences (Reconstruction, restoration, conservation, architecture)	*History* (Period system, revolution, event, phenomenon)

of human interactions, migrations, and ethnic identities across broad geographies [6, 7].

2.1 Dating Methods in Archaeology

The oldest and most used techniques in archaeology are undoubtedly dating methods. Carbon 14 can reveal the approximate history of artifacts along with dendrochronology and luminescence methods. Thermoluminescence reveals the date of the last heating of a ceramic object. It is possible to date most ceramics, clay objects, some rocks, flints, megalith structures, and even cave paintings [1, 8]. Thermoluminescence dating was first applied to a ceramic vessel fragment in the 1960s. A new method introduced in 2020 has enabled ceramics to be dated with Carbon 14 [9]. Archaeologists generally date ceramics stratigraphically and relatively. However, learning the exact date of certain questionable pieces will help to solve some problems. The obsidian hydration dating method developed in the 1960s is an aging analysis based on the measurement of water absorbed in obsidians, volcanic rocks. The Infrared-Photoacoustic Spectroscopy (IR-PAS) method has higher precision, correlating the water peak's height with rim width. Secondary Ion Mass Spectroscopy (SIMS) is the most accurate measurement with ± 0.005 μm error. Currently, the obsidian hydration dating method (SIMS-SS) introduced by scientists suggests the use of SIMS for the measurement of diffused water's profile in the outer surface layers of an obsidian tool [10]. Thus, it is used today as an exact dating method compatible with archaeological data [1, 11]. Another exact dating method is radioactive dating applied to bone and cave deposits [5].

2.2 Origin Determination and Characterization Studies

These studies aim to determine population mobility and commercial networks by identifying the source region of materials from various sources in close or remote regions. The materials analyzed in this context are both inorganic (faience, stone, ceramic, glass, sediment, pigment, metals, fossils) and organic (bone, skin, wood, textile, plant remains) [1]. These include obsidian source determination, mineral ore source determination, clay source determination in ceramics, and XRF (X-Ray Fluorescence), NAA (Neutron Activation Analysis), ICP-MS (Inductive Coupled Plasma-Mass Spectrometry), FTIR (Fourier Transform-Infrared Spectroscopy), and Raman Spectroscopy. The first important study on obsidian resource determination and dating emerged in the 1960s. Cann and Renfrew's *"The Characterization of Obsidian and Its Application to the Mediterranean Region"* (1964) and Friedman and Smith's *"A New Dating Method Using Obsidian"* (1960) marked a new era. To date, many methods have been added to various problems. Thus, regional syntheses and databases began to emerge.

Recently developed applications of ion mass spectroscopy and infrared-photoacoustic spectroscopy bring new methods for obsidian dating. Today, some archaeologists are still skeptical about obsidian hydration, but this will change with new technological advances [12]. A common use for the geological petrology technique is the examination of ceramic fabric under a microscope. The determination of rocks and minerals in additives to the fabric offers a comparison with specific geological sources in the region. This examination of fabric groups of ceramics can also be a reference for future studies by establishing a regional database. Also, it contributes to the determination of local or imported ceramics. Various periods have witnessed a detailed analysis of changes in ceramic production [13]. Lead isotope analysis (LIA) has been used in geological source determination of certain archaeological materials for more than 50 years. The determination of ore sources, particularly metal artifacts, as a result of chemical analyses is difficult due to the sources' similar geological features. A new study on this issue examines the change in Pb isotopic oxide and sulfide ore minerals. It correlates them with the geological histories that ore formation has in each region. In the areas where ore deposits are mostly similar in terms of geological ages, such as in the Andes Mountains, Europe, and the Mediterranean; it is inherently challenging to perform provenance analysis for Pb isotopes with similar isotopic ratios in sources of distant geographical origin.

Conversely, the regions with many ore formation periods, such as South Africa, are seen as quite promising for future studies done with Pb isotopes. As a result, resource analysis studies could produce better results in specific world regions. The number of analyses has increased following the introduction of MC-ICPMS in the late 1990s. This method works faster, cheaper, and more precisely than TIMS to measure Pb isotopic archaeological LIA ratios [14]. These studies have prompted research on determining archaeological materials and ore sources' chemical composition in recent years. Archaeometallurgy will continue to offer archaeologists a new perspective and evidence on the use of different resources and metal trade determination.

2.3 Archaeo-geophysical Methods

They are essential for determining artifacts or structures preserved under the ground using archaeo-geophysics before excavation. From the second half of the twentieth century, geophysical methods became widespread in archaeology. There are various scanning methods for this work. These include Ground Penetrating Radar (GPR), Electrical Resistance Prospection, and Magnetic Prospection-Magnetic Susceptibility [1, 15]. Accurate analysis of measurements using various scanning methods can provide important information about what is under the ground without any excavation. For instance, it is possible to identify an area with an architectural structure and its walls before excavation [15]. Geophysical studies can help determine the structure's depth, location, and shape without harming any buried structure is carried out in 3D. This is crucial in obtaining preliminary information about where to start excavation and the depth of buried structures [16]. It is even possible to scan the whole of an archaeological site with this method. A current study is the GPR and magnetometer survey study carried out in Falerii Novi, a Roman city in Italy. This study has uncovered the architectural plan of a Roman city without excavation. Archaeological excavation has provided detailed information about the city plan spread over a wide area and obtained over many years. Therefore, the concept of Roman city planning has generally been understood. The Roman city plan reflected a different architectural plan by contrast with more familiar towns such as Pompeii. In this study, computer-aided geophysical methods, which are not yet advanced, have been included.

Along with such new analytical methods, future studies will fundamentally change our knowledge, such as that not all Roman cities were designed like Pompeii, are expected [17]. Nowadays, different high-resolution remote sensing techniques serve to integrate into studies of landscape archaeology. They mostly comprise of a satellite (optical and radar data), air (photographic, infrared, and Lidar data obtained from aircraft and unmanned aerial vehicles), and land acquisition (different geophysical techniques, field walking, and differential GPS topographic research). These studies offer new approaches that have emerged in the last decade in geoarchaeology research [18]. Even though geophysical studies offer limited information to archaeologists, it is possible to achieve better results thanks to new scanning techniques emerging with each passing day. Today, it is not always possible to carry out scanning with geophysical methods for various reasons such as rough and mountainous terrain and conductive soil types. It is important to use the right method for the terrain conditions. New methods that are suitable for all types of land conditions are expected to emerge in the future. It is possible to compare results using more than one method before an archaeological excavation. Furthermore, the accuracy of data previously obtained with archaeological excavation is proven. Thus, one can expect a developing relationship between archaeology and geophysical sciences to positively influencing each other. Geophysics, helping archaeologists show what is unseen under the ground without any excavation, is the most vital aid to save time and budget by showing the right place to excavate. It is, therefore, among the indispensable instruments for archaeological field research. In the future, the geophysical methods will undoubtedly maintain their importance in archaeological studies.

2.4 Archaeoastronomy

Another archaeometry branch becoming increasingly important is archaeoastronomy. Prehistoric people determined some religious ceremonies, harvest seasons, calendars, and certain structures' locations by observing the sky and stars. It is a science that examines based on ancient astronomical observations in an archaeological structure or culture [1]. So, how did they manage to observe the sky? As a result of a long-term study of the sky with the naked eye, they determined that the moon, the sun, and stars moved at regular intervals. Even there were small stars invisible to the naked eye noted in Ancient Egypt. Archaeoastronomical studies determined that successful astronomical observations occurred mainly in Mesopotamia, Egypt, and the Maya. They used these observations in dividing time, determining seasons, and the location of various structures. Some Stone Age monuments in England, such as Stonehenge, are thought to have been built due to sky observations. It is still a mystery about how advanced ancient people were in sky observation. Future studies will help to develop new perspectives. The ancient Egyptian civilization applied knowledge of astronomy, mathematics, and geometry, especially for architectural structures. For example, a recent study tracked the application of geometry to a still earlier date in the Göbekli Tepe structures, approximately 11.000 BC. This study has solved one of the mysteries of constructing one of the oldest temples throughout human history. Surprisingly, it has revealed that geometry knowledge was in work as reflected in buildings constructed by planning (Fig. 2) [19]. This study shows surprising aspects of the distant past by understanding archaeological discoveries better with the positive sciences' support.

Fig. 2 General view from Göbekli Tepe (*Photo by Soner Atesogullari*)

2.5 DNA Archaeology

In 1985, archaeology began to examine biological remains. Bimolecular archaeology research helps archaeology obtain ancient DNA or aDNA (Paleogenomics) from skeletons and determine nutritional data (Isotopic Analysis). In recent years, DNA studies have revealed the origins of people who created cultures and the unknowns about migrations [20]. aDNA is based on separating the samples' DNA from archaeological bones, teeth, or mummified tissues by a special method. Today, the introduction of whole-genome sequencing methods from the cell nucleus has started to produce better results. An aDNA study examining the genetic structure of South-eastern Europe has revealed significant results. The origin of the farming that first appeared in Europe in the mid-seventh millennium BC is through immigrants settled in southeast Anatolia before moving to Europe. Understanding this process's dynamics was the objective behind the analysis of aDNA data of 225 people living in South-eastern Europe and surrounding regions between 12,000 and 500 BC. Previously, it was possible to identify the vast majority of European ancestry from three separate sources. The first one is ancestors associated with Mesolithic hunter-gatherer groups. The second is related to Neolithic farmers of North-western Anatolia and "North-Western-Anatolia-Neolithic-related" ancestry strongly associated with agriculture. The third is the ancestry emerging during Late Neolithic-Bronze Age in western Europe and associated with the "steppe" derived from Yamnaya steppe pastoralists. There is a reference to the steppe-related ancestry that involves a mixture of Upper Palaeolithic hunter-gatherers of Caucasia and the first farmers of northern Iran [7, 21]. However, there are still some questions that archaeologists cannot answer. When FOXP2 genes were discovered, this was assumed to be the genetic key for the difference between Neanderthals and humans. However, the main point of this study is not to understand human evolution. Reich and many researchers are skeptical that we can reveal biological or behavioral differences between our ancestors and us by comparing ancient and contemporary DNA. aDNA studies are expected to answer unknown questions about the origins of humanity. For instance, the determination of global human migration can give us information on many issues. Various developments are needed to produce the revolutionary ancient DNA evidence for archaeology promoted by Reich. S. M. Downes characterizes these in three categories: extraction methods, the quantity of sequenced genomes, and genetic analysis methods [22]. David Reich's contributions to aDNA studies are undeniable. The future of aDNA studies is quite exciting. With the help of developing new methods, Archaeologists can manage to answer new questions. aDNA has become one of the branches of science that helps archaeology today and will allow us to enthusiastically follow past evidence, such as tracing our ancestors' footsteps.

2.6 Conservation Sciences

It involves preserving and/or restoring artifacts or architecture by archaeological excavations. Cultural heritage management is one of the essential elements of archaeological excavation. It has two main functions, which are to present visuals for community archaeology and preserve archaeological materials. In a sense, the restoration and conservation of unearthed archaeological artifacts, especially architecture, help reduce excavation damage. The better preservation of architecture will undoubtedly allow visitors to an archaeological settlement to understand the settlement much more quickly. Restoration sciences that would enable visitors to complete restored walls with their imagination are significant for archaeology. The application of measures, such as preserving any archaeological finds in situ or moving them to a museum to implement preservation in a laboratory, is possible with restoration sciences. In conservation studies, the object should be documented before and after conservation by photography or video so that researchers and restoration scientists can follow the records [1, 13]. Architectural restoration and conservation in archaeological sites have provided the opportunity for archaeological site management and open-air museums.

For example, Turkey offers a unique open-air space in terms of archaeology. The open-air museum is necessary, especially whether monumental artifacts are as large as to be unmoved to a museum. Since the beginning of the last century, the preservation of artifacts in situ provided open-air museums' development. During the excavations started in 1946 at the site of Karatepe in the district of Kadirli in Osmaniye province in Turkey, in 1958, the "Karatepe-Aslantaş National Park" was declared under the direction of Halet Çambel and the "Karatepe Open Air Museum" was established [23]. The first open-air museum of Turkey was one of the pioneers in this field (Fig. 3). The restoration sciences are the most crucial help for archaeologists in raising a city from a small archaeological artifact. Innovations in chemical products and new methods and techniques can maintain cultural heritage conservation by allowing the restoration sciences to carry out more effective work.

2.7 Geoarchaeology

It emerged in the early 1970s that covers ancient environment and climate research with geology, geography, and environmental sciences. Modern archaeology integrates into a wider field of the natural sciences instead of being a branch of the arts and humanities by using geological methods [1]. Nowadays, geoarchaeology has become a multidisciplinary science that uses earth sciences in determining the human and physical environment in prehistoric times. One of the most important paleogeography studies in this field proves that today Troy's city was situated on the sea-coast during the Bronze Age in northwestern Anatolia. Sedimentological data obtained by drillings on the alluvial plains at the Troy have provided evidence to illuminate changes in the ancient environment. The shoreline's continuous change due to alluvial deposition of Karamenderes (Scamander) river deltaic

Fig. 3 "Karatepe-Aslantaş National Park" and "Karatepe Open Air Museum" at Kadirli town in the city of Osmaniye-Turkey (*Photo by Adem Yildiz*)

progradation greatly affected Troy city. After Bronze Age, Troy lost its location, harbor, and shore side. At present, Troy is situated on the Karamenderes plain nearly 7 km inside the seashore [24]. This study has shown how important it is to consider an archaeological site by assuming its environment. Paleo-environment tries to determine the environment at the time the archaeological site was in use. It is a multidisciplinary science that determines environmental conditions in the past by examining the physical environment, climate, land and sea conditions, soil structure, vegetation, and animals around the site together.

2.8 Cyber-Archaeology

It appeared as a discipline in archaeology in the early 2000s. It aims to animate archaeological data with virtual reality thanks to various computer graphics, such as digital simulations and 3D modeling. It is possible to make many different animations of the past by making wide use of computer science [25]. For instance, it is possible to raise a structure, including its all finds, in 3D. With new advances in the computer sciences and software engineering, cyber-archaeology applications are becoming more various every day. Modern computer software enables the use of statistical analyses that are complicated for archaeologists. The results obtained are only as good as the evidence from which they come from and the suitability of the testing methods used. A primary goal of statistical tests should be to provide shelter for archaeologists to collect and analyze data more accurately and precisely. The awareness of probability and correlation that will develop in time will undoubtedly offer a better understanding of sampling [13]. Information and Communication Technology (ICT) that has just started to apply for archaeology can store and use information. Currently, archaeology is developing research methods and techniques using ICT, the Internet, and Industry 4.0 information technologies.

2.9 Statistics

It is one of the sciences that help archaeology by solving specific problems with various analyses. These sciences help archaeology analyze the past using multiple data. Weaving statistics help us get information about the thread's thickness and woven fabrics' sizes. Thus, the production potential of hundreds of textile tools obtained in most sites can be calculated [26]. This has increased the amount of textile research in archaeology over the last two decades. The use of experimental archaeological methods in weaving has allowed many unknowns of archaeological textile production to be understood. Today, studies for developing new textile analyses and new methods continue. A recent study using statistics has tried to determine the spread of culture with histograms. The spread of the Neolithic to Europe has been a debated topic for many years. The spread of a culture can be revealed using mathematical modeling as a new archaeometric approach. It is possible to determine the cultural and demographic spread. For instance, a model was developed with histograms in a study carried out about the Neolithic spread to Europe. The spread rate, how much area it influenced, and points of finds were shown with graphs. This study indicates that statistics and mathematics can make raw archaeological information more understandable by presenting it graphically [27]. There is no doubt that mathematical modeling in immigration and culture spread will present important results in the future. The use of statistics and mathematics, especially with computer sciences, may allow new approaches and analyses to evaluate many archaeology data in the near future.

Archaeology, as integrated with natural sciences, includes:

- Archaeology is an interdisciplinary science that uses some methods of the natural sciences; and
- Natural science, whose methodologies ultimately involve many sciences (biochemistry, biology, chemistry, geography, geology, materials science, medicine, and physics), strongly correlates with archaeology.

Accordingly, when a new technique is discovered in the natural sciences, archaeological science can be quickly integrated into archaeological studies. For example, currently, aDNA analysis supplies new approaches to ancient human populations and migration theories.

3 Social Sciences and Archaeology

For the last 20 years, archaeology has benefited from economics, ethnography, history, sociology, philology, geography, philosophy, and anthropology to analyze a community in an archaeological site and the sociocultural adaptation strategies to the natural environment (Table 1). While archaeological finds and their analysis provide much information and are quite popular, research on the community structure has always been more limited. As a social science, archaeology can reveal past communities' social organization by using social sciences methods. For instance, archaeological and social sciences in Finland determined Neolithic hunter's social organization, gatherer fishers. The second phase focused on reanimating social networks' social structure and determination within and outside the household. Social organization analysis can determine the size, the number, and total population of settlements, what kind of a culture or community it was, how and why it changed, and what kind of natural and social relations existed between the culture/community and its environment. An analysis of the community can consider a limited amount of archaeological data [28]. New methods and theories in this field will do joint research in archaeology and the social sciences more common in the future. It is possible to obtain information about people and their environment thanks to analyses done using social science methods.

Ceramics, architecture, and stone tools are the main sources of archaeological data. At first glance, this may not seem ideal for analyzing social systems, economy/trade, or political dynamics. However, analytical methods and physical and chemical applications increasingly offer precise information on raw material sources and production. New computer analyses allow meta-analysis of these data. This allows the social sciences to contribute to archaeology. Social studies such as anthropology, economy, and trade encourage the study, especially for determining social structure and economic development and commerce networks [29]. Interdisciplinary studies of social, biological, and natural sciences play a very critical role. The social sciences and environmental sciences' joint studies determine human communities' nature and environment landscapes in a settlement through the ages [29]. Moreover, studies of daily life based on archaeological evidence can be analyzed using various social science methods.

It may be possible to establish a relationship between the social sciences, physical-engineering sciences, and archaeology in terms of archaeometry with meta-archaeology. Lester Embree considers archaeological study as a sub-discipline of the philosophy of science. He defines this as a conceptualization field with critical reflection rooted in archaeologists' dialogs, whom we can define as historians, art historians, museum scientists, restorers, etc. There are two basic dimensions in this field: while one of them is rational, historical, sociological, and economical, the other one can be evaluated in the framework of schools of thought increasing in the West, for instance, empiricist, evolutionist, feminist, historical materialist, neo-Marxist, technological, determinist, etc. [3]. The re-evaluation of material cultural remains along with the philosophy of archaeology, adds new dimensions in terms of interpretation. Evaluation of archaeometric results and archaeological knowledge is possible by evaluating the issue from a broad perspective from the philosophical point of view, in a meta-archaeological way [3].

Ethnography, perhaps one of the most important sciences to which archaeologists frequently refer, for them, is living prehistory. It is a multilayered science, just like archaeology. The ethnographic values alive today also include knowledge, techniques, or methods from the past. At this point, the examination of historical architecture, an oven, or a production technique allows a better understanding of archaeological finds. Ethnoarchaeology is a major, though it is not always the best way to provide operationally important analogy in explaining the archaeological records. Ethnoarchaeological data can also be exactly experienced through experimental archaeology. For instance, archaeologists can observe numerous details in the production phase of archaeological material. However, experimental archaeology cannot fully provide ethnoarchaeology's perspective due to its narrowness and lack of a cultural context [30]. Ethnoarchaeological data may give the best retrospective reconstruction for archaeologists. Sometimes people in the same region unwittingly use the same traditions and techniques for centuries. Therefore, ethnography carrying traces of the past to the present provides a new perspective for archaeologists in evaluating the past.

4 Archaeology and Society

Answers to questions such as "What does archaeology offer to a society? Why is it significant? Why should it be in interaction with society?" determine the role of archaeology as a social science. It is essential to stress the importance of archaeology to the individual who is the smallest component of society before its importance to society as a whole. Archaeology is a branch of science that illuminates humanity's history as a result of many years of study. Considering this framework, it does not directly affect the individual during his or her life. On the other hand, he/she can individually continue to progress in his/her life along with heritage awareness of his/her ancestors as a small representative of humanity. There is the help of known parts of human history, which will help him/her understand his or her location within the history of humanity when he/she looks back. A person developing himself/herself through archaeological science discoveries will undoubtedly step into the future equipped with knowledge of the past. Otherwise, he/she will not know where they come from and where to go and will finally become lost in terms of community. For instance, if an architect has mastered the architectural values of dwellings from the first examples in human history to the present, this will help him/her to reach unique horizons in terms of creativity, inspiration, and aesthetics. This example can touch many science fields, such as medicine, agriculture, ceramics, design, and technology. In short, as the study of everything in the past, archaeology should be one reference guide in an individual's personal development. When learning modern sciences today, archaeology sometimes answers questions that we did not know existed. In this sense, it is understood that archaeology is multi-layered and can inspire many areas by shedding light on the future. The importance of archaeology for society is that it is a science that enables society to advance more strongly in the future thanks to knowledge it obtains from the scientific study of the past. For example, archaeology, which has been carried out in Turkey for many years as an activity removed from society, has started to change with time thanks to many of the author's esteemed colleagues' efforts. By touching society, more in the future, archaeology will allow society to gain a broader perspective. We do not doubt that societies with knowledge revealed by archaeology will be more contemporary and more advanced in the future.

The role and importance of archaeology as a social science have long been debated. The effort of archaeologists studying such a long process as human history to return to the present and share their investigations with the present society is only one of archaeology's sociological aspects. From a more general perspective, as the science of the past, archaeology covers the universal aesthetic values of thousands of years ago, inspiring any humanmade artifacts such as modern homes, cars, jewellery, toys, parks, and monuments.

There is no doubt that all kinds of media are useful in archaeology, reaching today's communities. For instance, when the "Gypsy Girl" mosaic with her famously sad eyes was found during the Gaziantep Zeugma Excavation, archaeologists and the community as a whole touched not only archaeologists but also the community as a whole. Everyone found something of their own in the gaze of this gypsy girl looking at us with sadness from across the ages, as shown in the Art Performance. This was a crucial bridge established by archaeologists between the past and the future. Archaeology, the science of the past, seeks to illuminate our future by shedding light on human history's unknowns. In other words, it takes on the task of keeping the common memory of humanity by recording where humanity's adventure came from and where it is going from the Stone Age to the Space Age. The future can never be isolated from the past. The introduction of excavations and research done in archaeology, a human science, to the public using any media tools will allow new culture bridges. This assigns an important duty to archaeologists to carry out excavations and research and publish these results, besides explaining to society the importance of discoveries for humanity's history. Archaeoparks or archaeological site centers, of which various examples we have seen in world archaeology in recent years, are points where archaeology touches society just like experimental archaeology practices performed with society members.

Archaeology, as integrated with social sciences, includes:

- Archaeology is an interdisciplinary science that uses some methods of the social sciences; and
- Social science, whose methodologies ultimately cover a broad range of sciences, including economics, ethnography, history, sociology, philology, geography, philosophy, and anthropology, correlates with archaeology.

Archaeology mainly focuses on ancient human activity and its environment from a cultural perspective. As a social science, archaeology has the task of building a bridge between today's man and ancient man. In this context, it is important to share archaeological findings with today's society. Archaeopark projects in ancient settlements and education in museums constitute important focuses of archaeology's opening up to society.

5 Conclusion

Archaeometry, founded by physical and chemical methods, may have formed the first generation of archaeological sciences. Particularly between 1950 and 2000, basic archaeometric analyses started and began to be updated. From the 2000s to the present, new techniques and methods began to emerge in archaeometry. The second generation, which can also be called the evolution period, is continuing. An innovative third generation, which will continue with quite different analyses as shown by today's improvements, is expected in the future given current developments. Advances in the hard sciences (physics, chemistry, biology) are beginning to be integrated into problem-solving in archaeological field studies. Thus, archaeological sciences' contribution increases in understanding an archaeological settlement, an ancient community, or culture or reaching accurate and proven information.

In light of current studies, archaeology is unimaginable without the positive sciences. Developments in these branches of science will also affect archaeological studies. In other words, a positive correlation between the positive sciences and archaeology is on the agenda. New discoveries or devices will maintain their importance in achieving precise results in archaeological studies. While all these developments make archaeology an interdisciplinary science, they may also allow it to interact with all sciences. In this context, many joint studies, e.g., population dynamics, population distribution, etc., can be done with the social sciences. In the future, archaeological studies can develop new perspectives to the extent that it benefits from the positive sciences or social sciences in understanding the past. New scientific methods and discoveries in the archaeological sciences will undoubtedly open new horizons that we cannot even imagine today in the field of archaeology. Thus, the future of archaeology is quite exciting. Archaeologists attaching importance to interdisciplinary studies will illuminate the unknowns of the puzzle of humanity's history more decisively than ever before. In this responsibility framework, it is necessary for archaeologists to follow current developments, have basic information about which materials can be analyzed, and include them in their projects. Today, many archaeologists are unaware of archaeometric analysis and still trying to excavate using old methods. Future archaeologists are expected to be trained in archaeological sciences and to be better equipped. In this respect, the archaeologists who first uncover archaeological finds are like a conductor. Indeed, they are the key persons who will provide solutions to many archaeological problems by determining which analysis they need in any conditions, cooperating with relevant science branches, and developing projects.

Archaeology directs archaeometry. In other words, archaeological sciences help solve an archaeological problem, and a new study is carried out using new theories and methods they offer. This allows archaeometry to develop solutions and remain up to date continually. It is like a locomotive carrying the science of the past, archaeology, to the future. Today's archaeometry applications clearly show us that it will not be possible to carry out an archaeology project without archaeological sciences in the future. The use of new technologies integrated into archaeological science in the future will bring expansion in the field of archaeology that we cannot foresee today. Studies bringing together the social sciences, art, and archaeology are beginning to emerge.

Core Messages

- Archaeology is an interdisciplinary science involving natural and social sciences.
- Over the past decade, archaeological fieldwork underwent dramatic changes associated with archaeological prospection technologies and methodologies.
- Archaeological sciences and archaeometry play an important role in investigating our past and protecting buried and still standing heritage.
- Recently, new approaches with the help of analyses of social sciences and humanities have helped us understand ancient social structure, economy, settlement models, etc.
- When considering the possibilities that the first archaeologists had at their disposal 100 years ago, it would be possible to predict future archaeologists become equipped with a fertile source of fascinating technological or scientific possibilities that significantly facilitate information extraction from archaeological sites.

References

1. Liritzis I, Laskaris N, Vafiadou A, Karapanagiotis I, Volonakis P, Papageorgopoulou C, Bratitsi M (2020) Archaeometry: an overview. Scientific Cult 6(1):49–99
2. Türker L, Demirci Ş (1998) Şapinuva, Ortaköy (Çorum) Kazısından Elde Edilen Bir Tekstil Örneğinin Analizi. In: Arkeometri Sonuçları Toplantısı, pp 39–42
3. Demirci Ş, Balcı A (1998) Metaarkeoloji: Arkeometri Arkeoloji İlişkisi Üzerine. Arkeometri Sonuçları Toplantısı 1998:31–37
4. Libby W (1969) Radiocarbon dating. Chicago University Press, Chicago

5. von der Osten H (1937) The Alishar Hüyük seasons of 1930–32, vol Part III. vol OIP 30. Chicago
6. Utge J, Se´vêque N, Lartigot-Campin A-S, Testu A, Moigne A-M, Ve´zian Rea (2020) A mobile laboratory for ancient DNA analysis. PLoS ONE 15(3):1–21. https://doi.org/10.1371/journal.pone.0230496
7. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528(7583):499–503
8. Roberts RG (1997) Luminescence dating in archaeology: from origins to optical. Radiat Meas 27(5–6):819–892
9. Casanova E, Knowles TD, Bayliss A, Dunne J, Barański MZ, Denaire A, Lefranc P, di Lernia S, Roffet-Salque M, Smyth J (2020) Accurate compound-specific 14 C dating of archaeological pottery vessels. Nature 580(7804):506–510
10. Patel S, Hedges R, Kilner J (1998) Surface analysis of archaeological obsidian by SIMS. J Archaeol Sci 25(10):1047–1054
11. Liritzis I (2006) Sims-SS, a new obsidian hydration dating method: analysis and theoretical principles. Archaeometry 48(3):533–547
12. Current Questions and New Directions in Archaeological Obsidian Studies (2016) Oxford University Online. Accessed July 11, 2020
13. Greene K, Moore T (2010) Archaeology: an introduction. Taylor & Francis Group, Florence
14. Killick D, Stephens J, Fenn TR (2020) Geological constraints on the use of lead isotopes for provenance in archaeometallurgy. Archaeometry
15. Smekalova TN, Yatsishina EB, Garipov AS et al (2016) Natural science methods in field archaeology, with the case study of Crimea. Crystallogr Rep 61(4):533–542
16. Zorlu K (2017) Determination of Buried archaeological structures by archeogeophysical methods (Olba, Mersin, Turkey). Cumhuriyet Sci J 38(4):184–192
17. Verdonck L, Launaro A, Vermeulen F, Millett M (2020) Ground-penetrating radar survey at Falerii Novi: a new approach to the study of Roman cities. Antiquity 94(375):705–723
18. Deiana R, Leucci G, Martorana R (2018) New perspectives on geophysics for archaeology: a special issue. Surv Geophys 39(6):1035–1038
19. Haklay G, Gopher A (2020) Geometry and architectural planning at Göbekli Tepe Turkey. Cambridge Archaeol J 30(2):343–357
20. Brown TA, Brown K (2011) Biomolecular archaeology: an introduction. Wiley
21. Roodenberg SA (2018) Biyolojik Antropoloji ve Arkeolojide Kullanılan Yöntemler. In: Ünlüsoy S, Çakırlar C, Çilingiroğlu Ç (eds) Arkeolojide Temel Yöntemler. Ege Yayınları, İstanbul, pp 271–314
22. Downes SM (2019) The role of ancient DNA research in archaeology. Topoi, pp 1–9
23. Çambel H (2010) Karatepe-Aslantaş Açık Hava Müzesi: Çok Yönlü Bir Proje. TÜBAKED 8:131–136
24. Kayan İ (2018) Jeoarkeoloji ve Paleocoğrafya Araştırmalarının Arkeolojideki Yeri. In: Ünlüsoy S, Çakırlar C, Çilingiroğlu Ç (eds) Arkeolojide Temel Yöntemler. Ege Yayınları, İstanbul pp 17–63
25. Jones IW, Levy TE (2018) Cyber-archaeology and grand narratives: where do we currently stand? In: Cyber-archaeology and grand narratives. Springer, pp 1–17
26. Yılmaz D (2016) Textile production in Western Anatolia in the late bronze age: loom weights from Çine-Tepecik Höyük. TÜBA-AR Turkish Acad Sci J Archaeol 19:93–113
27. Fort J, Pareta MM (2020) Long-distance dispersal effects and Neolithic waves of advance. J Archaeol Sci 119:105148
28. Kıvımäkı S (2008) Archaeology and the social sciences studying the social organization of neolithic hunter-fisher-gatherers in Finland. Suomen Antropologi J Finnish Anthropol Soc 33 (2):95–103

29. Smith ME, Feinman GM, Drennan RD, Earle T, Morris I (2012) Archaeology as a social science. Proc Natl Acad Sci 109(20):7617–7621
30. Politis GG (2016) The role and place of ethnoarchaeology in current archaeological debate. World Archaeol 48(5):705–709

Derya Yilmaz is Associate Professor at Ankara University, Turkey. She received her Ph.D. at Ankara University in 2010. She has held as a visiting scholar at the University of Cambridge, Freie Universität Berlin, Eurasia Department of German Archaeologisches Institute, Berlin. Her specialization is Early Bronze Age Archaeology in Anatolia and the Ancient Near East, focusing on cultural interactions. She has worked as a team member at Alaca Höyük, Acem Höyük, Köşk Höyük, Liman Tepe, Çeşme Bağlararası, Sirkeli Höyük, and Resuloğlu Höyük. She was co-director of Maydos Kilisetepe Excavation in Eceabat-Gallipoli Çanakkale between 2012 and 2017. She was director of Çanakkale central province archaeological field survey Project from the Neolithic Age to the end of the Iron Age conducted in years between 2016 and 2018. Currently, Derya is the director of the Külhöyük excavation at Gölbaşı-Ankara, which the Ankara Anatolian Civilizations Museum excavated for 14 seasons between 1992 and 2006 years.

22 Biomimicry in Architecture: From Theory to Practice

Mehdi Azizkhani

In all things of nature there is something of the marvelous.

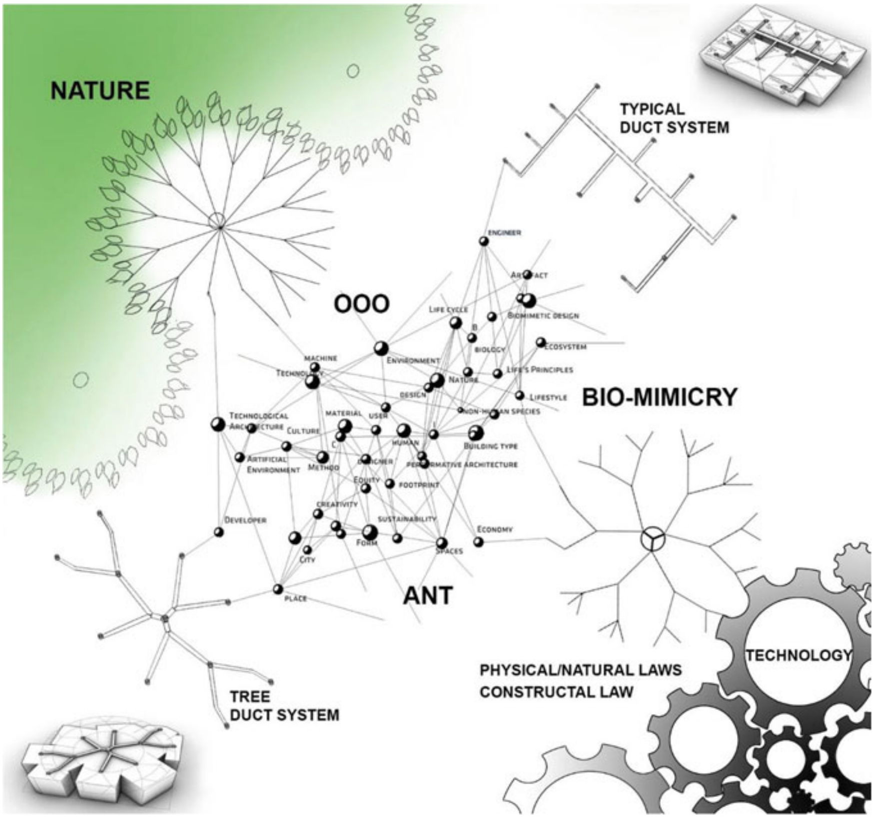
Aristotle

Summary

Nature and technology define two key components of design approaches. Nature has inspired us since the time of our existence to live with it, while technology can help us today to live even against the harshest natural conditions. On this basis, one camp of designers idolizes the role of nature in design as in biomimicry, while the other camp stands by technological solutions to address our design problems through technology. This chapter will address a balanced approach focused on building design/construction between the two groups' approaches by addressing the flaws and potentials of their approaches in theory and practice. A test case focused on tree structures, duct designs, and building layouts will be introduced and analyzed to see the application of the proposed approach. The findings suggest that there is currently a vague connection between the conceptual findings in nature and their actual implementation in building design and construction. More in-depth scientific analysis and investigation of the laws behind natural phenomena are required to fill this gap, as demonstrated by the test case.

M. Azizkhani (✉)
Texas A&M University System, College Station, TX 77843, USA
e-mail: m.azizkhani@tamu.edu

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_22



Connecting the domains of theory and practice between nature and technology focused on a test case of the tree duct system.

The code of this chapter is 01000001 01110010 01100011 01101000 01101001 01110100 01100101 01100011 01110100 01110101 01110010 01100101 00001101 00001010.

1 Introduction

Nature and technology are two fundamental notions of design approaches. The debate over the applicability of the biological or technological design approaches [1], with their deficiencies and potentials in shaping our built environment, has formed two extremes of subversive and subservient in advocating nature or technology. Figure 1 shows key examples of contributions, events, and discussions regarding technological and biological concepts in chronological order. In some cases, the contribution of nature in a design process is idolized at the cost of

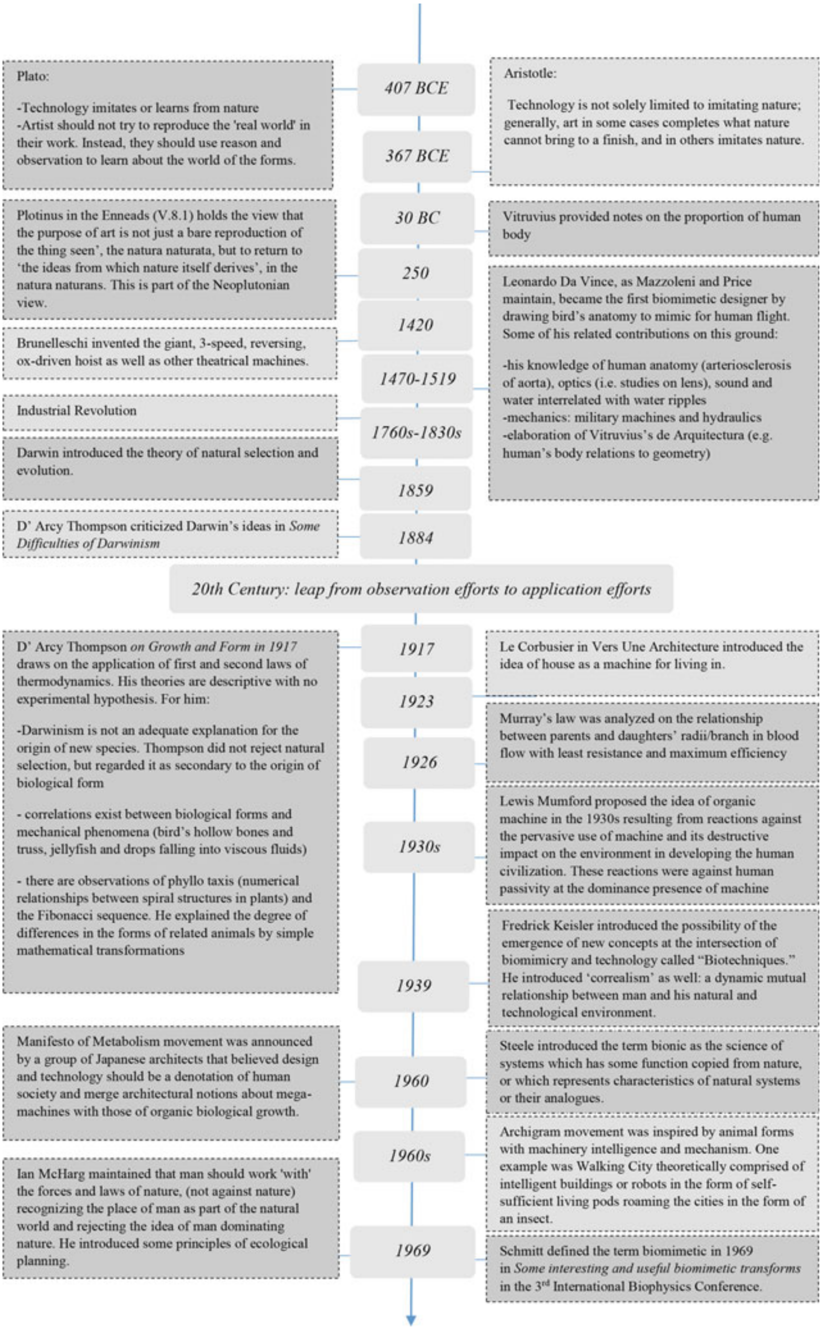


Fig. 1 Examples of biological and technological contributions and concepts in chronological order

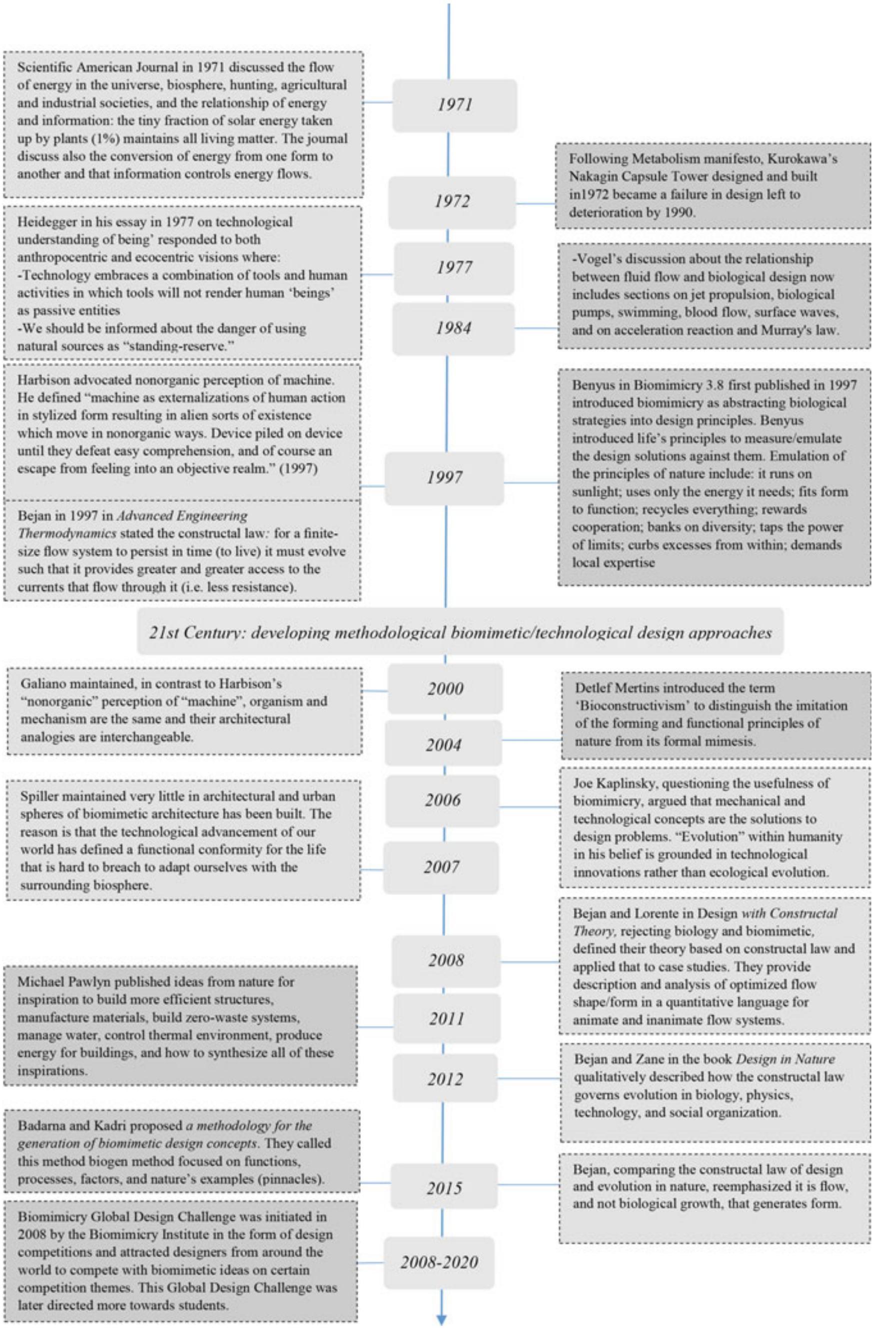


Fig. 1 (continued)

ignoring the associated technological achievements. In some other cases, our design creativity is restricted because of imposed technological restrictions to the point of repetitive boredom and human passivity.

This chapter tries to address a balanced approach standing in between the two extremes. It compares biomimicry, defined as the "abstraction of good design from nature" and "the conscious emulation of nature's genius" [2], with the technological approach that represents applied engineering [3]. The former designs a building like nature would build, perusing continuity, while without pursuing continual evolution, the latter may interfere with or transcend the natural design by connecting the parts to form a distinctive structure [4, 5]. Both approaches have unique features that can be combined to develop a more comprehensive method for designing improved building envelopes and systems.

However, it seems that attempts to connect these two domains in architecture have not been properly addressed. Although technological solutions in practice have been responsive to human needs for centuries, the new designs resulting from biomimetic approaches in architecture have rarely reached an implementation phase, which is understandable from several cases that we will review. Yet, beneath the surface of many of today's technologies are biologically similar systems and principles. These principles and the natural laws behind them can be discovered and developed through appropriate approaches in building design and construction to be integrated with other technological advancements. Therefore, there is a need to address two questions: first, why have biomimetic approaches rarely been implemented in the architecture and construction industry? Second, how can we fill this gap to address the integration of biomimetic and technological domains in architecture and construction to inspire building envelope and building systems by nature? In this chapter, first, I maintain that a more practical approach in the design of building components lies at the intersection of biomimetic and technological solutions rather than one of the two domains. Second, through a test case, I try to demonstrate an approach to designing building components inspired by learnings from nature, where both technology and biomimetic domains overlap. Such an approach would not only be guided by the laws of physics and technological tools but also would incorporate biological processes, where appropriate, to produce an improved design.

2 The Fallacy of Biomimicry

Some industry leaders predict that between now and 2050, biomimicry will become one of the main tools in our move from an industrial era to an ecological era [6]. For example, a quick review of the US National Aeronautics and Space Administrative (NASA) and Science Journal websites helps substantiate this claim because each website shows de facto and potential biomimetic achievements. Examples are numerous, ranging from highly efficient LEDs inspired by fireflies to the strong metal materials inspired by skeleton bones and the bioinspired exploration systems emulating aerial fliers.[1]

However, very few biologically inspired architecture projects have been built [7], reflecting a gap in the architectural design world where biological concepts rarely turn into built functional forms [8]. A critical review/synthesis of the previous studies in biomimetic architecture outlined in the following section reveals that despite any theoretical connection, there is no well-accepted, practical method to connect the findings from an example at the scale of nature to the application at the scale of a building or its components. In contrast, the prevailing belief is that technological solutions rather than biological solutions provide the most effective response for our design problems. In other words, biological languages/analogies that idolize nature might downgrade or ignore designers' real achievements today [3]. Not considering combinations of technology and biomimicry as an integrated approach could be one reason: first, for such a loose connection between nature-inspired concepts and their practical application; and second, for the desirability of technological solutions over the biomimetic solutions.

Rarely, the existing biomimetic methods that were reviewed—in the next section—have stepped beyond their current form to reach a concrete product on the scale of a building component. Most biomimicry publications in building design are concluded in conceptual tables and diagrams for classification, conceptual development, or analogies with nature to be used for design concept production. These diagrams, while useful, are not going beyond conceptual levels and are insufficient for the implementation of our learnings and findings in nature. There is often a missing step at the end of the existing methodologies to connect the findings of a biological process to a final design product through a visibly logical procedure. This gap could be the result of the lack of a more in-depth and detailed approach in architecture to learn from existing natural processes to be integrated with technological tools for robust architectural productions inspired by both nature and technology.

Biomimicry has been investigated in the literature at different levels, such as natural forms, processes, or ecosystems. Here, the author focuses on the biomimetic "process" and "form" for inspiring building designs. The focus on natural forms for inspiring building design should not be interpreted as a metaphorical or literal translation of natural forms because theoretically and philosophically, neither architecture nor nature can be reduced to each other. Therefore, for example, mimicking the bodily form of a fish as a building plan to fit different rooms of a building is not what we pursue by discussing the use of natural forms in design: a fish is a fish and a building is a building and not reducible to each other. Lack of attention to this theoretical understanding might be one reason for the failure of

[1] For example, see the following: http://www.techbriefs.com; http://science.sciencemag.org; http://www.techbriefs.com/component/content/article/lighting-technology/lighting-technology/16495; http://science.sciencemag.org/content/355/6329/1055; and NASA's Jet Propulsion Laboratory http://www.techbriefs.com/component/content/article/ntb/tech-briefs/mechanics-and-machinery/961.

many biomimetic architectural concepts in reaching a successful implementation. The present chapter brings actor–network theory (ANT) and object-oriented ontology (OOO) into the context to substantiate this theoretical claim.

In the 1980s, ANT emerged to explain a foundational *theory of social research in technology*, which has also been applied in various fields such as environmental design. ANT assigns equal agency to different actors and the context defined among the actors with an alliance, relation, and translation [9]. Actors indicate human/non-human agents, which form through their relations in a network. ANT assumes that there is nothing outside the relationship network, there are only enacted alliances, and nature, technology, animals, or human/non-human agents are not different in their ability to act. Therefore, when an actor is engaged with an actor–network, it means the actor is also engaged with the web of relations [10, 11].

ANT mainly focuses on the relationships between things. However, OOO notices the things per se and their existence at all scales from a molecule to a behavioral construct in a society. In OOO, an object is not depending on other objects or their relations. This independence also includes the human perception of the object, where the object withdraws itself from being exhausted by perceptions. Likewise, all relations may induce the distortion of related objects, while all objects position themselves as equally important simultaneously [12]. Combining ANT and OOO can also form new theoretical constructs for design. For instance, Kärrholm, through re-conceptualizing "building types," has developed the concept of "territorial sorts" to examine buildings' roles in daily life [13].

Applying both OOO and ANT to our discussion on nature and technology, we can perceive that: first, within a certain assemblage of things/objects, our access to objects are only the descriptions that came from their actions. This access is always an incomplete and partial understanding because objects can interact differently in a different network/assemblage of things. Second, we perceive that nature, technology, architecture, form, function, and many other things are objects that withdraw themselves from being the subject of one single interpretation by actors, such as clients, users, or designers, in a building design/construction process. Therefore, the production of objects cannot be only through their relations, such as biological analogies, which use relational approaches. This will ignore the reality of objects and cause a shift in the reality from being formed of the designed objects per se to having something more "real" than objects. Although real objects exist in the built and natural environments, seeing an object's entireness in all forms of appearance and using it in all possible ways seems impossible [13]. Thus, while every object holds autonomy, its effect hinges on the object's relationships with other objects: an effect different from that of the objects in pieces *where no object is reducible to another object(s)* in the assemblages of objects [14]. As a result of this autonomy, for example, a biological growth function is irreducible to the growth of the Nakagin Capsule Tower, or, for example, a giant self-sufficient walking machine is irreducible to a self-sufficient mobile city. The first example is representative of what architects were pursuing in the metabolism movement in the 1960s and the 1970s, and the second example was one of the ideas pursued by architects in the Archigram movement almost around the same time. History has shown the failure of such movements when such ideas later became obsolete or were never realized [3]. To further clarify the inefficiency of such bioinspired ideas and irreducibility of objects in nature to building/building components, we will review several themes in the following section. This review will also help us to find the gap in the translation of our findings in nature to practical solutions for our building designs and constructions.

3 A Brief Overview: Biomimetic Trends and the Practical Gap

This section will review three themes focused on biomimicry to understand better its current trends and the existing gap in its application in buildings. These themes include:

i. A brief review of the existing biomimetic methodologies [15–21];
ii. A review of an example of biomimetic design for a building component through the above approaches; and
iii. A brief analysis of the mechanism of a cooling system utilized in the East Gate Mixed-use Center in Zimbabwe, Africa, which claims to have been a successful bioinspired design imitating the function of termite mounds [22].

The importance of the first two areas of this review is that they identify the often ineffectiveness of the current biomimetic approaches in developing building envelope and system components in architectural design. The third area, adding to the previous two, indicates how we may emulate findings in nature to reach a successful building/building component design.

a) In general, existing biomimetic design approaches are divided into two categories of problem-based and solution-based approaches used by different groups who study biomimicry [16]. Solution-based approaches lead to a design product in which the observation of nature has inspired the design process. In contrast, problem-based approaches are the approaches for a specific engineering problem in which a solution is being sought from nature. On this basis, different categories of biomimetic approaches can be identified. For example, some of the scholars have identified several groups of biomimicry advocates who have used solution-based and problem-based approaches [16]. Four of these groups include Group1. *Biomimicry 3.8* including Biomimicry 3.8 [15] and Benyus [2]; Group2. *BioTriz* including Vincent et al. [20, 21]; Group3. *Biomimetics for Innovation and Design Laboratory* managed by Shu and Vakili [18, 19]; and Group4. *Design and Intelligence Laboratory* led by Goel [17]. While all of these groups have drawn on the problem-based approach, some (e.g., Group1 and Group4) have also utilized the solution-based approach.

The general steps of the problem-based approaches include 1. definition of the problem; 2. exploration and investigation of the natural models for solving that problem; and 3. developing solution ideas based on the findings in nature. For example, Group1 and Group2 approaches can follow these steps: 1. defining the problem/challenge to be solved; 2. finding the related functions in nature responding to this challenge; 3. developing an analogy between nature and the required functions in reality; 4. abstracting the functions found in nature and extracting their identified principles to be used for design concept generation; 5. developing a design concept by emulating these principles and evaluating them against life's principles. Examples of nature's life principles include running on sunlight, using only the energy you need, fitting form to function, recycling everything, and relying on diversity [2]. The solution-based approach follows the same steps of the problem-based approach, only in reverse: (i) biological domain investigation for finding a natural system, model, or solution; (ii) transferring the findings into solutions and design principles; and (iii) trying to connect to the technological domain by defining the correspondent problem, emulating nature's design principles, and testing/prototyping solutions.

Scholars take issues with these biomimetic approaches due to their difficulties in translating nature's functions to functions appropriate for architectural design [16]. Therefore, they also propose another methodology called BioGen to generate biomimetic design solutions. Accordingly, the BioGen methodology follows six steps to reach its outcome:

- Defining the design challenge;
- Discussing possible scenarios and finding an organism/system in nature that forms an ideal adaptation strategy (i.e., pinnacles);
- Analyzing selected pinnacles;
- Deriving imaginary pinnacles (i.e., the most appropriate pinnacles related to the required functions for mimicking);
- Outlining the design concept, which is the superposition of the imaginary pinnacles to determine the dominant features to be addressed in the next step; and
- Generating a preliminary design concept.

Although BioGen seems to be a comprehensive design tool and, according to the authors proposing the BioGen approach, easier to be used for biomimetic design, its outcome is still conceptual and challenging to be converted into a real product. The authors themselves mention this shortage: "…the main limitation of the BioGen methodology is that it does not provide a transition from the concept phase to the emulation phase" [16]. This shortage of the trending biomimetic design approaches, including the most recent BioGen method, indicates the literature gap that the current methods and approaches of biomimicry are very general, leaving the result at a rough conceptualization level.

b) A review of the application of these methods in architecture reinforces this flaw and their methodological inefficiency in practice. For example, several studies employing the BioGen method have focused on optimizing the building envelope's performance. However, they seem too simplistic and idealistic to reach the implementation phase in the building industry. To analyze one of these studies in detail, I came across an example of biomimicry using the above methods to learn from nature for building envelope thermoregulation. After discussing a performance taxonomy of organisms for thermoregulation, this case demonstrated a possible application to building envelopes through an assembly of bioinspired components. The main component of this assembly was the stoma brick made of a porous material, which was supposed to mimic the human eyelids at the outer layer to remove the dust from the air passing through the envelope. Inspired by the pinecone skin, a veneer shutter would control the opening and closing of the envelope modules according to a humidity gradient. The inner layer of the stoma brick would use a HEPA filter for air cleaning, and the innermost spongy layer was to hold moisture for evaporation in dry climates or absorption in humid climates. An irrigation cycle in dry climates would be added to the system to irrigate the bricks. This system was designed to operate in different climates: for example, in hot, humid weather, the veneer shutter would deform due to increased humidity, which will let the air stream move to the inside of the building while passing through the spongy layer. Although this is a creative design concept, in reality, the stoma brick system may not necessarily function as it was described here, and its actual performance remains a question.

Clearly, the impact of humidity on this system has been downplayed. In the most idealistic condition, when the spongy component absorbs the moisture in the air at some point, it will be saturated with 100% relative humidity. After this saturated condition, no more absorption would occur, leading to the formation of water droplets on the walls' surfaces due to increased relative humidity: the result will be an indoor environment with too much moisture as well as deterioration of the building envelope due to moisture accumulation. Additionally, the HEPA filter performance depends on the key factor of face velocity and a large pressure difference to force air through the filter. Therefore, in conditions with normal wind speed, it may not effectively remove the unwanted particles from the incoming air.

This example epitomizes one of the flaws traceable to other similar biomimetic designs: proposing inefficient envelope components because of a literal translation of a design concept to a physical product following the existing biomimetic design methods. Unfortunately, these methods do not link their conceptual products, at the last step, to the resulting building system/envelope component through a visibly systematic and scientifically grounded process.

c) In one example that shows promise as an application of biomimicry, the cooling system of the Eastgate Mixed-Use Center in Harare, Zimbabwe, cools down the outside air to save on energy consumption by 90% compared with a conventional air conditioning system. It is claimed that termite mounds have inspired this building's cooling system [23]. The reason is that a large termite mound includes air pockets, elaborate ventilation holes, and air passages, which effectively trigger natural ventilation through convection. This process keeps the termite mound temperature between 84.2 and 89.6 °F in a hostile climate with a fluctuating temperature of 35 °F at night to 104 °F during the day. Two models for termite mound ventilation are commonly accepted: thermosiphon flow driven by heat and induced flow driven by wind [24].

In the case of capped mounds, based on the thermosiphon flow model, the hot air built up in the nest rises to the mound top to be evaporatively cooled with water vapor through the porous walls of the mound. Then this denser, moist air descends to the nest below in a repeatable cycle. The induced flow describes the mechanism of the mounds with a chimney on the top and uses the stack effect, where the chimney helps to accelerate the buoyancy-driven flow. The unidirectional flow pulls in the fresh air from near the ground to the nest, which will be passed on through the chimney and to the outside at the end. The Eastgate Center building drawing on both buoyancy-driven and wind-driven flow models, moves the air inside the building through its walls and floors' expansive tube/duct systems. Internal heat gains along with the heat stored in the building's structure trigger a thermosiphon-effect to draw up the air toward the rooftop with large chimneys. The chimneys also help to create an induced flow.

Although termite mounds inspired part of the concept behind the Eastgate Center's design, which is highly admired as a biomimetic design, other mechanisms, such as its air conditioning system, which draw on machinery mechanism and technology, have been ignored in the existing analyses. For example, the building utilizes fans with a low speed during the daytime and fans with a high-speed during nighttime to prevent the formation of stagnant air. This process can effectively replace the built-up hot air in daytime hours with cool air at nighttime hours. Smaller fans run during daytime hours to keep the indoor environment comfortable while the walls gain and accumulate the outdoor environment's heat. Larger fans during the nighttime run to pull out the accumulated heat from the walls and push it out through the wall/ceiling ducts. The walls will be ready to release the stored cooling again by the next morning.

Although the combination of passive and active systems here saves on using expensive air conditioning technology, clearly, no termite mound utilizes fans [22]. Therefore, this building, in the absence of a machinery contribution, could not perform effectively and cannot materialize its biomimetic design concept. As a result, what we seek in a design inspired by nature is not usually a literal translation of a form or phenomenon found in nature; it is more the physical laws/processes we try to learn or analyze to inspire our designs. In some cases, such as emulating a termite mound's function to design a natural ventilation/cooling system in the Eastgate Center, we draw on the physical principles of these natural processes to inspire our designs. In contrast, in some other cases, we use machinery and our technological advancements to act against these natural processes to provide human comfort, as in the case of providing thermal comfort inside our homes. The use of low speed and high-speed fans in Eastgate Center was one of these cases where we may look at nature to mitigate its unsuitable conditions to make a harsh climate comfortable for living. Therefore, instead of looking at nature through the lens of biomimicry alone, we need to draw on ANT and a wider network of actors that beyond nature includes, for example, technological advancements, machines, the laws of physics, users, and many other factors along with their interrelations as well as a wider network of professions such as architects, engineers, physicists, simulation professionals, biologists, or in one world an interdisciplinary network of professions to trans-discipline the final outcome.

4 Filling the Gap

As we briefly reviewed, very few scholarly works have attempted to trace buildings or building components to biomimetic processes, or vice versa, by a detailed and scientific study of the physical laws governing the natural processes. This deficiency might hint toward a better understanding of why technological solutions have been responsive to human needs, while the results produced from biomimetic approaches in architecture have rarely reached the implementation phase or scientifically reasonable conclusions. Different engineering disciplines have used the laws of physics and experimental tools to fill this gap [25]. A case in point is the set of recent optimizations in thermodynamics that focus on generating optimized geometric forms in animate/inanimate flow systems [26].

In this case, Sieniutycz and Jeżowski consider limits on both the energy and size of organisms in biosystems and demonstrate how physical laws are significant in creating organisms' optimal forms. In their analysis of thermodynamic optimizations, they draw on Bejan's Constructal theory by focusing on the number of bifurcations of the human respiratory tree structure and emphasize the function of the physical laws in animate/inanimate systems. In these systems, the optimization process is searching for a geometric form with the minimum flow resistance regarding time, volume, and weight constraints. The resulting optimized structure is called "constructal design," and the law governing design development is called "constructal law" [27]. In the constructal world, nature uses thermodynamic principles (i.e., first and second thermodynamics laws) to guide the design process to produce optimal designs [28].

According to the second law of thermodynamics, a tree structure's self-organization happens in nature because the entropy of a child branch resulting from bifurcation is less than its parent branch. In other words, biological systems are allowed to evolve a bifurcation tree structure in a hierarchical mode because the combined entropy of both child branches is higher or equal to the parent branch. Therefore, in Bejan's constructal law, the final goal of evolution is established by the second law of thermodynamics' move toward randomness in nature, which governs modification/speciation [26].

Although the constructal law has not been applied to architectural design, it has been applied in different fields of engineering, such as in the case of heat transfer distribution in power plants, flow channel and fin size/shape optimization, and aerodynamic/hydrodynamic shape optimization to maximize flow and minimize fluid friction. The following section, by drawing on the constructal law of energy/nature, simulation/modeling tools, and the science and art of architecture, will depict an example of learning from nature to inspire a building component design.

5 A Test Case in Inspiring Design by Nature

In response to the gap mentioned above in the literature (i.e., translating a concept found in nature to a building component), here I draw on the constructal law of nature to illustrate a methodology to inspire the design of a building component and its form while integrating the laws of nature with modeling and simulation tools in an early design phase. The constructal law of nature [29] will govern this process in reaching the desired form.

We may think living organisms have a random form, as described in the branching patterns of a tree. However, their forms are, in fact, the result of a continuous process of performance optimization. Studies on tree networks and their branching configurations have shown in many cases they can perform better than manufactured structures due to the forms of their optimized conduits with a lower flow resistance [30]. Studies in this regard started as early as the 1920s by Cecil Murray' study on the blood vessel system [31, 32]. Murray's law demonstrated an optimum relationship between the vessels' diameters in a human body to carry the blood with minimum work and flow resistance. This relationship has been shown in the following equation between the diameter of the parent vessel (D_k) and its two offspring branches (D_{k+1}) [33]:

$$D_{k+1}/D_k = 2^{-1/3}$$

Per Bejan and Lorente [27], the tree network configurations in nature can reach optimized forms due to specific optimization, but not random, reasons: *For a finite-size system to persist in time, it must evolve such that it provides easier access (i.e., minimum flow resistance with respect to global constraints of time, weight, and space) to the imposed currents that flow through it.* Accordingly, the three significant factors in the performance of a tree structure include length, angle, and level (i.e., step) of branching. There is not a fixed law for the optimized length of tree-shaped conduits,[2] and the optimal length ratio is usually determined based on constraints such as area and volume in addition to flow resistance [32]. The 75°

[2] Except for the one level of T-shaped single branching, in which the ideal length ratio is as same as the diameter ratio.

angle between symmetrical branches is the ideal angle for an optimized performance, which should be modified at each step of branching to ensure the final tree structure stays within its defined boundary. As shown in Fig. 2, this angle could be 120° depending on the distance between the perimeter nodes of a tree configuration [27].

After investigations of the physical laws and principles of an optimized tree structure in nature, these principles were applied to a typical office building. The goal was to promote its duct configuration and, accordingly, its room layout. Accordingly, I defined five rules/principles to reconfigure the office plan and its duct system:

- Hierarchy to assign larger rooms to the first level of branching nodes for a higher air volume flow rate and smaller rooms to the next branching levels;
- Adjacency to minimize distances in the new rooms' layout for neighboring rooms compared with the original plan;
- Symmetry and circularity to maximize the symmetrical and circular form of the tree within the existing building footprint;
- Bifurcation angle of 75° or close to 75° for outer circles and 120° for the inner circle to optimize the tree structure of the flow system; and
- Murray's law to determine the sequential branches' diameters according to the equation mentioned above.

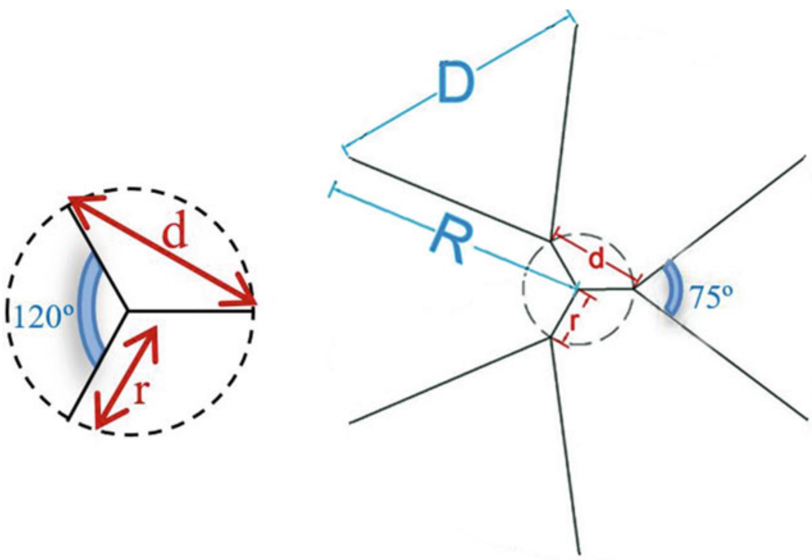


Fig. 2 The best tree form when D is smaller than, but comparable with R, and when d is larger than r

The effort was to make the new room layout stay within the existing building footprint, but if a portion of the new layout went outside the building footprint, it was cut out and used to fill the gaps between rooms inside the building footprint. The same scenario applies to the overlapped areas of the rooms to be merged or cut to fill the building footprint. Based on the five rules above, a tree structure with a minimum flow resistance, which has one branching level, was useful to define the air distribution network in a typical office building. The office building size determined the size and dimensions of the tree structure, but if needed, this structure can expand according to Bejan's findings for optimized tree structures [27]. The assumption was that air disruption location is at the center of each individual room in the existing building, which uses a usual right-angle duct system (Fig. 3).

Similarly, for the proposed room layout, the effort was to introduce air into each room at its center. The existing office building layout was modified by realigning and overlaying each room's center on the appropriate node in the proposed tree structure (Fig. 4). Therefore, the nodes on the tree structure represent the air diffusers, and the tree structure makes the basis of the proposed tree duct system. The Air Handler Unit (AHU) is at the center of both the existing and proposed plan layout to reduce the ducts' length and the resulting static pressure drop. The resulting proposed room layout (Fig. 4) can also be redefined by applying the genetic algorithm (GA) in rhino/grasshopper software to produce more genes or, in other words, more room layout alternatives. One example of this application can be

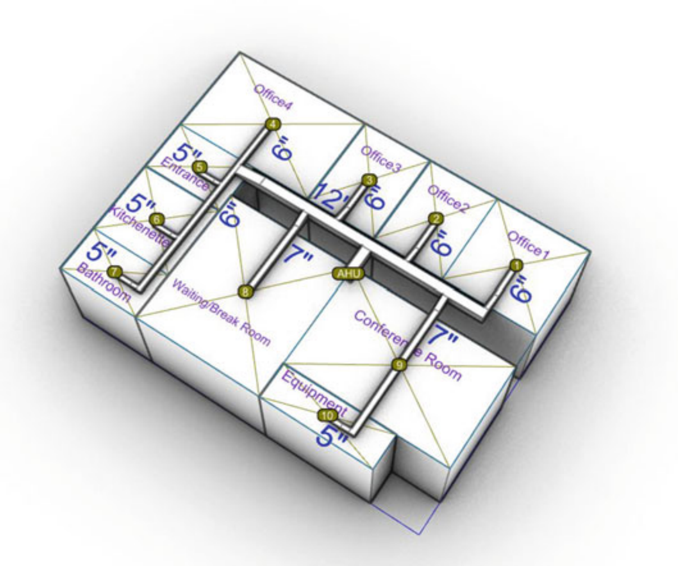


Fig. 3 The office building used as a test case in its existing condition along with duct sizes

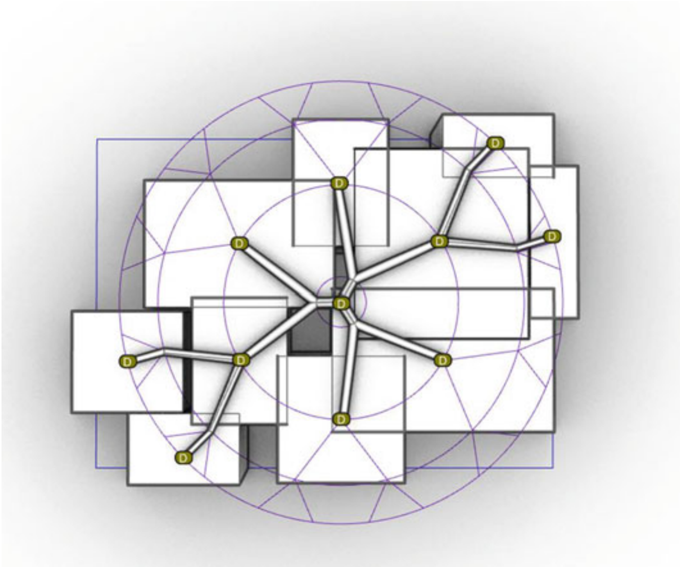


Fig. 4 The office layout after realigning its rooms' centers with the tree structure nodes according to the five rules

found in Fig. 5. While not the focus of this study, GA here represents another example of how nature can inspire building design.[3]

As we see, the tree duct system in Fig. 4 or Fig. 5 is inspired by the tree configuration rules. The round duct size of 9″ at level zero was determined using duct charts [34], and the diameter was reduced gradually based on Murray's Law to 8″, 7″, and 6″ in connected branches. To evaluate the performance of the tree duct configuration versus the common right-angle duct system, I conducted a CFD simulation focused on one comparable branch of each duct system. Table 1 shows the simulation features for the one selected branch of each of the two duct systems, including the tree duct system and the right-angle duct system. This branch starts from the AHU at the center of the office and extends to the farthest air distribution point.

The CFD simulation in the selected branch of the tree duct system showed a better performance than that in the usual right-angle duct system. The tree duct system had a static pressure drop of 34–37% when approaching the air outlet, which is lower than that of the right-angle duct system with a 55% pressure drop. Additionally, the smooth turns in the tree duct system branch showed less turbulence at

[3] A genetic algorithm is a metaheuristic search in computer science inspired by the process of natural selection in Charles Darwin's theory on natural evolution. In this algorithm, the next-generation offsprings are reproduced by the selected fittest individuals for reproduction. By defining a rhino/grasshopper graph for optimization through the Galapagos component, a genetic algorithm was used in this study to find the proper orientation for each room inside the original building footprint following several constrains including the five rules defined in this section.

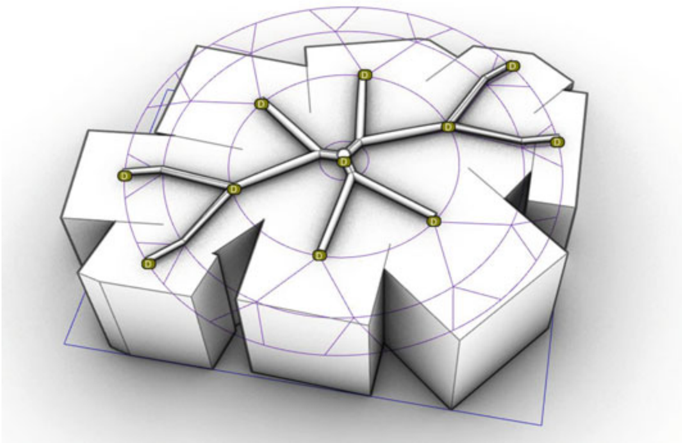


Fig. 5 One of the five optimized genes using grasshopper and GA along with the superimposed conceptual tree duct system

each turn as another advantage of this system compared with the usual duct system. In other words, the static pressure at the corners in the tree-duct branch dropped 20–30% less than the right-angle duct branch, thereby taking less time after each turn for the flow to gain a uniform profile. Figure 6 shows the results of the simulation for each duct branch regarding the air velocity profile comparison.

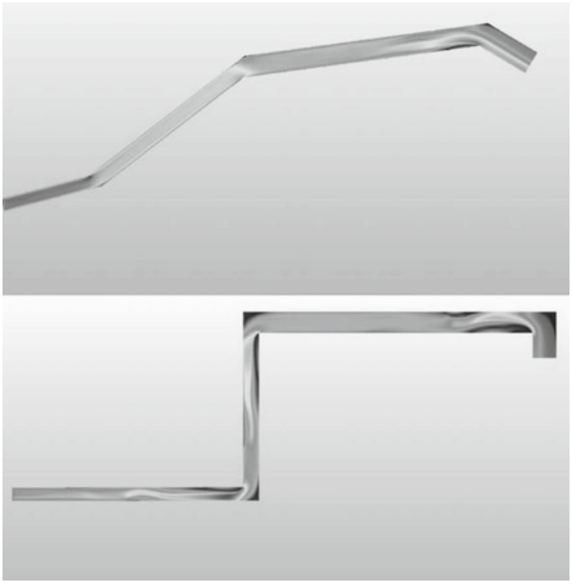
6 Conclusion

It has been said that there is something of the marvelous in nature. However, bringing the marvelous findings of nature from the level of thought to the level of practice needs a comprehensive, transdisciplinary vision and approach. As we discussed, except for very few cases, biomimetic examples of building designs have not successfully reached the implementation phase. Classification diagrams and conceptual tables for biological analogies or conceptual diagrams for emulating our findings in nature are useful; however, they usually leave us with our findings at the concept level and are insufficient to direct us to the construction and implementation of design ideas. Therefore, there is a need to draw on other actors and disciplines to extract the scientific laws behind natural processes to inspire our designs. This is achievable in a network of actors as in ANT, including, for example, various architectural design disciplines, engineers, simulation professionals, physicists, clients, users, biologists, and many other experts where the final outcome is transdisciplinary.

Table 1 Simulation features for the two duct systems

Simulation	Duct System	Duct length	Mesh size	Environmental conditions	CFD model	Advection	Iteration	Boundary conditions
1	Usual right angle plenum system	22.6’	0.2	100 kPa 20 °C	SST K-Omega	Advection 5/modified Petrov–Galerkin	3000	Inlet: Gage zero static pressure Outlet: volume flow rate of 22 cfm
2	Proposed tree structure system	18’	0.2					

Fig. 6 CFD simulation results showing the change of air velocity in one branch of each duct system



Considering both OOO and ANT, in a certain assemblage of objects, our access to objects is only a description of their actions. This is always incomplete access and a partial understanding because objects can interact differently in a different assemblage. Meanwhile, nature, architecture, technology, form, building components, and many other things are objects that withdraw themselves from subjection to one singular interpretation, including that of a solely biomimetic design. Therefore, objects cannot be produced only by their relationships, as in the biological analogies that use relational approaches. The reasonable approach in nature-inspired designs, when appropriate, would not only incorporate biological processes but also would be guided by the laws of physics and technological tools to produce an improved design. This seems to be a current gap hindering the translation of what we may find in nature to a useable building design/product that can be implemented. Therefore, we reviewed a test case focused on tree structures in nature to see how we can fill this gap.

As we saw in tree structures, the constructal law of nature is one of the physical laws of nature that can help us connect nature and technology to inspire our building designs. By investigating this law and extracting the principles governing the optimized performance of natural tree structures, we established five rules to redesign the duct system and room layout of an existing office building. In our redesigning process, we applied different modeling and simulation tools, and we applied what we saw in nature through scientific investigations to inspire the design of the proposed duct system and room layout. We tested and compared the common type of duct system currently applied in buildings with the proposed tree duct system to show the better performance of the tree duct system. While here we have focused on the flow of fluids and its impact on new room layouts, we should not ignore the importance of other design factors such as visual, thermal, and spatial comfort. In other words, for the ease of our case study/analysis, we focused on one item here (i.e., flow), but we need to incorporate other design factors and modify our designs to reach a more comprehensive and all-inclusive design solution at the end.

Nature has always been a source of inspiration for us since the time of our existence. It provided us with creative ideas and concepts. In the past, man was able to mimic simply the concepts visible in nature, such as living in prehistoric caves mimicking animals living in dens. However, today, after a long time of building upon our findings in nature and advancing technological achievements, we are considerably beyond that point. We cannot go back as we have tasted the comfort of a technological lifestyle with all of its advantages and disadvantages. Instead, today, we need to search for the right approach to connect and develop what we can find in nature to potentially and practically applicable strategies and products for our current and future lifestyles. This connection can take shape if we deeply and scientifically investigate our discoveries in nature by a detailed analysis of their governing natural laws to be either emulated for new technological achievements or to act against them to be able to live in harsh natural conditions when needed. These scientific investigations, analyses, and emulations require our transdisciplinary perception and collaboration from the level of findings in nature to the level of application and implementation.

Core Messages

- Neither architecture is nature nor is nature reducible to architecture.
- Existing biomimetic trends mostly provide a vague, conceptual connection between findings in nature and their actual implementation in buildings.
- ANT and OOO may help us develop a transdisciplinary theory and vision to adopt the right approach between biomimetic and technological design solutions.
- Many fields and actors, beyond biology, architecture, and engineering, should be involved to facilitate this transdisciplinary concept-to-product materialization.
- Scientific and detailed investigation of natural laws through technological means, such as laws of energy and simulations in tree structures, may facilitate this process.

References

1. Cohen P-S, Naginski E (eds) (2014) The return of nature: sustaining architecture in the face of sustainability, Routledge
2. Benyus J-M (2002) Biomimicry: innovation inspired by nature. William Morrow, New York
3. Kaplinsky J (2006) Biomimicry versus Humanism. Wiley, Great Britain

4. Braham W, Hale J (2006) Rethinking technology: a reader in architectural theory
5. Keisler F (1939) On correalism and biotechnique: a definition and test of a new approach to building design. Architect Record 60–75
6. Pawlyn M (2011) Biomimicry in architecture. RIBA Publishing, London
7. Mazzoleni I, Price S (2013) Architecture follows nature: biomimetic principles for innovative design. CRC Press, Taylor & Francis Group, Boca Raton, Biomimetics series
8. Spiller N (2007) Good-natured stuff. Archit Des 77(6):144–145. https://doi.org/10.1002/ad.589
9. Latour B (2005) Reassembling the social: an introduction to actor-network-theory. Oxford University Press, Buckingham
10. MacKenzie D-A, Wajcman J (1999) The social shaping of technology, 2nd ed. Open University Press
11. Williams R (2019) The social shaping of technology (SST). In: Pittinsky TL (ed) Science, technology, and society: new perspectives and directions. Cambridge University Press, Cambridge, pp 138–162. https://doi.org/10.1017/9781316691489.006
12. Harman G (2005) Guerrilla metaphysics: phenomenology and the carpentry of things. Graham Harman. Open Court, Peru
13. Kärrholm M (2013) Building type production and everyday life: rethinking building types through actor-network theory and object-oriented philosophy. Environ Plan D Soc Space 31 (6):1109–1124. https://doi.org/10.1068/d15312
14. Harman G (2010) Time, space, essence, and eidos: a new theory of causation. Cosmos Hist 6 (1):1–17
15. Biomimicry 3.8. http://biomimicry.net. Accessed 7 Sept 2020
16. Badarnah L, Kadri U (2015) A methodology for the generation of biomimetic design concepts. Archit Sci Rev 58(2):120–133. https://doi.org/10.1080/00038628.2014.922458
17. Goel A-K, Rugaber S, Vattam S (2009) Structure, behavior, and function of complex systems: the structure, behavior, and function modeling language. Academic Press, Great Britain
18. Shu L-H (2010) A natural-language approach to biomimetic design. Academic Press, Great Britain
19. Vakili V, Shu L-H (2001) Towards biomimetic concept generation. In: ASME design technical conferences, design theory and methodology conference, Pittsburgh, PA
20. Vincent J, Bogatyreva O, Pahl A-K, Bogatyrev N, Bowyer A (2005) Putting biology into TRIZ: a database of biological effects. Creativity Innov Manage 14:66–72. https://doi.org/10.1111/j.1476-8691.2005.00326.x
21. Vincent JFV, Bogatyreva OA, Bogatyrev NR, Bowyer A, Pahl A-K (2006) Biomimetics: its practice and theory. J R Soc Interf 3(9):471–482
22. Turner S (2008) Beyond biomimicry: What termites can tell us about realizing the living building. In: Proceedings of 1st international conference on industrialized, intelligent construction
23. Pawlyn M (2011) Michael Pawlyn on biomimicry. ArchDaily. http://www.archdaily.com/185128/interview-michael-pawlyn-on-biomimicry. Accessed Apr 2016
24. Haitham A-H (2011) Investigation of flow through and around the Macrotermes michaelseni termite mound skin. Loughborough University Institutional Repository. https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/8466. Accessed June 2020
25. Collins MW, Atherton MA, Bryant JA (2005) Nature and design. International series on design and nature: [1]. WIT Press, Southampton, Boston
26. Sieniutycz S, Jeżowski J (2013) Energy optimization in process systems and fuel cells, 2nd edn. Elsevier, Kidlington, Oxford
27. Bejan A, Lorente S (2008) Design with constructal theory. Wiley, Hoboken, New Jersey
28. Bejan A (2016) Life and evolution as physics. Commun Integr Biol 9(3), e1172159. https://doi.org/10.1080/19420889.2016.1172159
29. Bejan A, Zane JP (2012) Design in nature: how the constructal law governs evolution in biology, physics, technology, and social organization, 1st ed. Doubleday

30. Kou J, Chen Y, Zhou X, Lu H, Wu F, Fan J (2014) Optimal structure of tree-like branching networks for fluid flow. Elsevier Science B.V., Amsterdam., Netherlands
31. Murray CD (1926) The physiological principle of minimum work: the vascular system and the cost of blood volume. Natl Acad Sci United States of America 3:207–214
32. Sherman TF (1981) On connecting large vessels to small. The meaning of Murray's law. J General Physiol 78(4):431–453
33. Xu P, Wang XQ, Mujumdar AS, Yap C, Yu BM (2009) Thermal characteristics of tree-shaped microchannel nets with/without loops. Elsevier Science B.V., Amsterdam, France
34. Grondzik WT, Kwok AG (2019) Mechanical and electrical equipment for buildings. Wiley

Mehdi Azizkhani is Adjunct Faculty at Texas A&M University System and a Researcher at Texas A&M Energy Systems Laboratory. He is an ACSA, IBPSA, and ASHRAE member and holds several degrees in environmental design, including his Ph. D. in Architecture from Texas A&M University. Dr. Azizkhani has designed and managed different residential, commercial, and urban projects. Enriched with interdisciplinary professional/ academic experience in building design and construction, his area of work and interest addresses both the theoretical and practical aspects of environmental design, including sustainable design, nature-inspired design, passive/active systems, energy efficiency, building information modeling (BIM), building energy modeling (BEM), parametric design, design thinking, and architectural theory. Dr. Azizkhani has been the recipient of several national and international awards focused on building design and energy research, such as Building Performance Simulation Association (IBPSA), Texas Energy Summits, Texas A&M University, Texas Architectural Foundation, and Australia and New Zealand Architectural Science Association.

Ecodemocracy in Practice: Examples of Forestry and Food Production

23

Helen Kopnina, Reingard Spannring, Marco Morini, William Lynn, and Francisco J. Santiago-Ávila

Man is the only creature that consumes without producing. He does not give milk, he does not lay eggs, he is too weak to pull the plough, he cannot run fast enough to catch rabbits. Yet he is lord of all the animals. He sets them to work, he gives back to them the bare minimum that will prevent them from starving, and the rest he keeps for himself.

George Orwell

Summary

Some argue that instrumental anthropocentric values are not the motive enough to protect species and habitats perceived as functionally useless for human

H. Kopnina (✉)
The Hague University of Applied Sciences, International Business, Johanna Westerdijkplein 75, 2521 EN The Hague, The Netherlands
e-mail: helen.kopnina@northumbria.ac.uk

R. Spannring
Leopold Franzens University Innsbruck, Innsbruck, Austria
e-mail: reingard.spannring@uibk.ac.at

M. Morini
Dipartimento di Comunicazione e Ricerca Sociale, Universita' degli Studi di Roma "La Sapienza", Rome, Italy
e-mail: marco.morini@uniroma1.it

W. Lynn
George Perkins Marsh Institute, Clark University, Worcester, MA, USA
e-mail: wlynn@clarku.edu

F. J. Santiago-Ávila
University of Wisconsin, Madison, WI, USA

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_23

welfare. In the current democracies, there is little recognition of "intrinsic value," "rights," or "interests" of individual nonhumans and nature as an entity. This chapter aims to address this gap by inquiring what the practical implications of ecodemocracy using examples of forestry and food production are. This chapter addresses if existing democracies were to operationalize ecocentrism in policy-making, what would food production and forestry look like? Moreover, what are alternatives to feeding and housing billions of people without treating "land" just as crops and treating forest, not just as construction wood, decorative hedges, or recreational spaces? This chapter raises practical questions about when ecodemocracy is applied in real-life situations and discusses ways forward by addressing these questions. Some existing forms of democratic representation of nonhumans, or suggestions for possible forms, are discussed in this chapter, including the application of the precautionary principle, the grassroots organizations such as Council of All Beings or the Parliament of Things, the Parties for Animals, and still a possible one, the proxies.

One poster says "*Together Back to Normal" (protest against anti-COVID-19 measures), while the other poster says "We cannot go back to normal, because normal was precisely the problem.*"

The code of this chapter is 01000101 01100011 01101111 01100100 01100101 01101101 01101111 01100011 01110010 01100001 01100011 01111001.

1 Introduction

Many integrated environmental science researchers have wondered whether current democratic systems can adequately address environmental challenges, such as climate change and biodiversity loss [1–6]. Not only do most modern societies and politics appear human-interest centered [7, 8], but anthropocentrism also increasingly affects decision making in biological conservation [9–11] within global organizations addressing biodiversity loss. For example, the World Bank links healthy ecosystems to *"services that have significant economic value in many cases*" [12]. Another example comes from the most global democratic institution, the United Nations, which has developed 17 Sustainable Development Goals (SDGs). The first two goals are ending poverty and hunger, with key strategies outlined in section Food and Agriculture, emphasizing the importance of "sustainable *use* (sic) of the environment," "natural resources," "ecosystem services," "sustainable use of biodiversity" (https://sustainabledevelopment.un.org/). SDG 15 speaks of investment in forestry, emphasizing *"people and their livelihoods, especially the rural poor, youth, and women"* (https://sustainabledevelopment.un.org/topics/forests). Such anthropocentric (human-centered) terminology in SDGs [13, 14] is significant because these aims act as an implicit global blueprint and broadly translate into many countries' national policies concerned with environmental protection.

Some argue that instrumental anthropocentric values are not the motive enough to protect "left-over" (functionally useless for human welfare) species and habitats [10, 11, 15–17]. Treating biodiversity or nonhumans as "natural resources" or "ecosystem services" embed the "reigning belief that the Earth is our property: humanity's commonwealth" [18].

In the current democracies, there is little recognition of "intrinsic value," "rights," or at least broad "interests" of individual nonhumans and nature as an entity [8, 19, 20]. There is a complex relationship between democracy and biodiversity. However, recent research shows a disconnect between the urgency of issues and implementation. As Rydén et al. state [21], the main issue is that biodiversity produces responses following human disturbances, and these responses occur nearly instantaneously (e.g., hunting or deforestation) or in extremely long term (e.g., competitive exclusion by invasive species, population recovery). Other than time, another challenge is that at present, *"only loss in human life and welfare"* is seen as the basis of political action, without taking *"into account the consequences for nonhuman species and plants"* [4]. Some suggested that "*if there are no institutional guarantees other species will be considered in decision-making processes, then their interests will be constantly given low priority or completely ignored"* [3, 22]. Another issue is the policymakers' skepticism toward scientists and scientific evidence [23].

Following this, an anthropocentric ethic underlying broader democratic processes seems to condone actions that, directly or indirectly, could send more than half of terrestrial species extinct, for example, due to habitat loss [24, 25], leading to "biological annihilation" [26]. Simultaneously, as the human population and material demands increase, billions of animals are subordinated in the intensive food production system [18]. The land use for food production differs from Leopold's [27] seminal concept of "land ethic." Leopold proposed that the boundaries of the land's stakeholder community need to include collectives such as soil and species, such as plants and animals. This context should not only put species but all habitats and the entire biotic community to moral rights. Leopold's argumentation was able to balance all actors and dynamics involved. In this context, a rights discourse is not dependent on economic or utilitarian approaches to value nature [28]. However, it lies in deep ecology [29] and Earth Jurisprudence [30].

Some existing forms of democratic representation of nonhumans, or suggestions for possible forms, will be discussed below. Some of them distinguish between individual animals' rights and the rights of collectives (species, biodiversity, and habitats). In line with (eco)feminist thought, none of these needs to be categorical or a priori, but situated (i.e., contextual, circumstantial, and presumptive), with an acknowledgment of what is incalculable in any ethical–political decision [31, 32]. Context informs relationships and interdependency, and thus what is special or more salient to those involved. Ethical meaning arises from particular situations, as ecofeminists have proposed and promoted, for both individuals and collectives. In that sense, any theoretical ethical ranking could be anthropocentric to the extent that it prizes theory (logos) over situatedness. Likewise, it prizes independent qualities (autonomy) over relations (interrelatedness) or vulnerability [31, 33]. Species ranking has long been debated in both deep ecology and animal rights literature [34]. Compounding the potential of being mired in open-ended discussions of which individual animals or habits should have more or less ethical standing, nature rights overall remained challenging to implement [17].

This chapter aims to address this shortcoming. For this, it will focus on two specific research questions: If existing democracies were to operationalize ecocentrism in policy-making, what would food production and forestry look like? What are alternatives to feeding and housing billions of people without treating "land" just as crops and treating forest, not just as construction wood, decorative hedges, or recreational spaces?

We aim to address this gap in this chapter by inquiring what the practical implications of ecodemocracy using examples of forestry and food production are. It also aims to raise practical questions when ecodemocracy is applied in real-life situations and discuss ways forward. The sections below will discuss the four forms of representation, involving existing ones, e.g., the precautionary principle, the grassroots organizations such as Council of All Beings or the Parliament of Things, the Parties for Animals, and still a possible one, the proxies.

Below are definitions used in this chapter:

- Ecodemocracy refers to "*political processes that recognize the intrinsic value of nonhuman nature*" through inclusive pluralism [35]. Ecodemocracy sometimes holds in close association with terms "ecojustice" and "Earth jurisprudence," which is a form of democracy in which *"political decisions can be* *reached that account for the intrinsic value of nonhumans"* by establishing legal levers to discourage environmental harm;

- Agriculture refers to the science of cultivating plants and livestock for human consumption;
- Forestry refers to the science and craft of creating, managing, and using woodlands and associated resources for human and environmental benefits;
- Nonhumans or animals refer to both vertebrates and invertebrates and other living entities such as plants and bacteria;
- Nature refers to the environment, ecosystems, or habitats that might have been altered by humans;
- Rights (of species) refers to a set of legal and/or political and moral recognition of interests of a species, based on the recognition of "intrinsic value." While "animal rights" protect an animal from being oppressed or dominated by humans, "animal welfare" might permit the instrumental use of animals as long as they are used "humanely." "Nature rights" might refer to ecology-centered ("ecocentric") perspective and protection of "nature." By contrast, in "anthropocentrism" (human-centeredness), the environment is protected for the sake of human welfare;
- Intrinsic value refers to "*a value that is independent of any benefits for humans.*" In nature, it occupies a place in the center of the ecosystem to "ecocentrism" (ecosystem-centeredness). The concept of ecocentrism is related to "deep ecology," which is characterized by advocacy for a radical restructuring of modern human societies following the recognition of basic rights for nature. These rights can apply to individuals within the species, whole species, or even whole habitats, such as "land" or forests; and
- Ecorepresentation and proxies refer to humans who represent nonhumans or nature through ecodemocracy. Generally, groups and social movements that support ecojustice, ecodemocracy, or the fight against ecocide support the mutual flourishing of human and nonhuman species.

2 The Possibilities of Democratic Representation of Nonhumans

Perhaps the largest democratic organization dealing with nonhumans' rights is the Harmony with Nature program of the United Nations (http://www.harmonywithnatureun.org/rightsOfNature/), which supports educational activities on the rights of nature in the professional and public spheres to advance Earth Jurisprudence worldwide. The program explicitly aims to establish a new, nonanthropocentric paradigm, which serves as a more ethical basis for Earth and human relationships. However, this program seems to work independently of national democratic systems. Alongside UNESCO's declarations on the rights of nature and animals, the Earth Charter's moral principles seek to inculcate respect for the entire community of life [36–38]. It stands in contrast to the anthropocentric tendencies in education for SDGs [14].

The constitutional entrenchment of the precautionary principle in the functioning of more extended democratic systems has been proposed [39]. The precautionary principle emphasizes caution, pausing, and review before leaping into innovations that may prove disastrous, for example, in the case of human action affecting biodiversity [40–42]. The extended use of this principle would ensure that *"we would be gearing up towards a genuinely ecologically sustainable economy and society"* [39]. The precautionary principle applies when activity potentially threatens the health or the environment, recommending interventions to avoid potential negative consequences even before they are fully established scientifically, which could take time [42, 43]. For example, in the coronavirus outbreak at the time of writing, some governments have taken the precautionary principle regarding illegal wildlife trade based on emerging—but not yet solidified—scientific evidence that wild species consumption such as bats or pangolins caused COVID-19 [44].

Another possibility of giving a voice to nature does not involve political representation but civil group activities, like The Council of All Beings, developed in 1988 by Joanna Macy [45] and John Seed [46]. The Council activity involves the organization of workshops and rituals, where "*participants step aside from their human identity and speak on behalf of another being*." This could be a nonhuman individual or an entity expressing that being's concerns and wellbeing and the surrounding environment. Consequently, the participants talk as humans about their responsibilities to remove the threats or correct some of the injustices they have identified. The Council of All Beings method is normally not applicable to a political context. In contrast, it intends for the audiences wanting to learn more about "*their place in the ecosphere as humans and wishing to further develop an emotionally richer response to Earth's destruction*" [8].

Based on Latour's concept of "entanglements" [47], and the Parliament of Things (https://theparliamentofthings.org/about/), a group of Dutch creative entrepreneurs, established the North Sea Embassy, taking as a starting point that the sea owns itself (https://www.ambassadevandenoordzee.nl/english/). The Embassy, consisting of "designers, policymakers, biologists, artists, lawyers, philosophers, and writers," researches how various sea-related "stakeholders" (living beings and nonliving objects connected to marine environment) "can become full-fledged members of society."

However, the problematic term is "stakeholder," given it seems to do more exclusion regarding who gets to be considered/decide/vote. There may be considerations of who is more vulnerable or whose claims are more urgent, but "stakeholder" may homogenize these as well, giving all equal vote/power regardless of their situation. Based on one of the author's participant observations attending Embassy's workshops in 2019, the participants consider "things" such as phytoplankton and codfish but also oil platforms and the fishermen as equally important stakeholders. While one might negatively affect the other through pollution and overfishing, "listening to the North Sea" remains impartial. It is in line with Latour's work, which has critics as it turns attention away from environmental problems toward technology and scientific and social processes, so that “things” can express their identities [48]. As Whiteside states about Latour’s work, the Parliament of Things undermines the traditional justification of representative democracy without offering convincing ethical limits, seeing all “things” as equal, and refusing to address environmental problems [48]. Another limitation of such “Councils,” “Embassies,” or “Parliaments” is that the participants are already “converted,” highly conscious individuals. However, potentially, these informal organizations can help empower individuals, including those from disadvantaged communities, through educational activities such as role play. It is critical to emphasize their role as they can extend grassroots participation. Such workshops do not require academic background or a high level of awareness and can happen as playful events to encourage broader participation.

Broader and more socially diverse are members of various animal rights and animal welfare organizations, such as the Parties for Animals, present in over a dozen countries [19, 20, 49]. Animal Politics EU, a group of animal protection parties, is represented by voters at the national level in Belgium, Cyprus, Germany, Netherlands, Portugal, Spain, Sweden, the United Kingdom, and the European Parliament [49]. Often, the focus of Animal Parties is the welfare of individual (often domesticated) animals, and they are often single-issue parties. Only recently, some of these political organizations have developed their manifestos further, and, like conventional Green parties, they have extended their focus on broader issues of sustainability and social justice.

Last but not least, ecorepresentation through proxies was proposed [2, 7, 8, 35]. Ecorepresentation involves an advocacy mandate, with *“representatives being appointed for the specific task”* [4]. While the proxies could operate similarly to Parties for Animals, expanding their focus from domestic animals and local conditions to wild collectives and ecosystems, proxy representation’s practicalities are not yet established. Parliamentary places could be reserved for proxies as a form of continuous affirmative action as a correction because nonhumans cannot represent themselves and have historically been excluded from representation. Affirmative action, which is typically applied to policies and practices preventing discrimination of one race, gender, sexual orientation, creed, color, and national origin, from another, could be expanded to nonhumans.

3 Ecocentric Food Production

In the case of agricultural production, a moral accent often exists on “feeding the world” with maximizing utility of “land” [50]. Critics have noted that intensive agriculture turns land into toxic monocultures [51] and enslaves millions of animals, institutionalizing industrial-scale meat production [18]. Despite the growing media attention and a recent consistent awareness for the issue, [52] figures on farm animals indicate annual increases in animals’ global consumption (https://www.animalequality.net/food). While the intensification of farming “feeds the world,” monocultural plantations compromise ecosystem integrity, and animal farming causes methane emissions [50, 52]. Also, this intensification is often associated with demonstrably negative effects on animal welfare [50]. Garnett et al. note that the aim of "feeding the people" [50] should be contingent upon acceptable standards of welfare as well as ecological sustainability [53, 54].

A vegan/vegetarian/omnivore critique complements the ecological one in this regard, while some vegetarians, for example, from traditional Hindu communities, might focus on small-scale animal husbandry as sustainable and ethical. Vegan support for these activities may be contingent on the context and the necessity of animal use. For example, if one lives in a developed country and has no need for animal products, even small-scale animal husbandry may be challenged as unjust. What these food ethics viewpoints share include an abiding critique of confined animal feeding operations (CAFOs) and an opposition to the enormous increase in animal agriculture forecast by the FAO. Also, industrialized agriculture failures regarding personal and public health, pollution, animal welfare, social justice, and ecological/climate sustainability play a role in this context, along with the toxic overconsumption of meat in westernized economies and cultures.

Even the near-term future is unlikely to ensure a healthy and well-fed global human population, while ecosystems are degrading [53–56]. While invasive species are often "blamed" for biodiversity loss, the massive habitat conversion into animal and crop agriculture, the so-called "ecological imperialism," is invisible as a *bona fide*, masked by "feeding the world" euphemism and thus impervious to the framing of "invasive alien species." Significantly, "feeding the world," SDG2, applies to one species only. Against this anthropocentric perspective, considering the environmental costs of intensive agriculture, organic community-based agriculture [51] has been advocated. However, even this type of agriculture has its drawbacks, given the scale required by growing food for billions of consumers. While more benevolent in terms of toxic outputs, organic and regenerative agriculture still requires a large (sometimes larger than for genetically modified crops) territory and irrigation [50]. If nonhumans' needs are considered, then voluntary noncoercive reduction of the global human population is obviously of importance [56, 57]. An oxymoronic "have your cake and eat it too" of the SDGs (keeping both a large and wealthy population and environmental integrity intact) is impossible without radical reduction of population and change of diet. For example, Calarco [58] argues for the need to "*push back against the dominant anthropocentric worldview that disregards other possibilities for animal life besides being meat for human consumption.*" Calarco also notes that deep ecology questions the very basis of animal use or consumption, arguing in support of ethical veganism.

The Dutch Party for Animals leaves the question of food open to consumers. Still, it does provide their electorate with information regarding the negative effects of meat consumption on animals, nature, and humans [19, 20] (Fig. 1).

Fig. 1 Climate protesters from party for animals, March 10, 2019 (*Photo by Helen Kopnina*)

Even in cases where deliberative democracy cannot resolve conflicts, it can allow us to hear silenced voices [59]. The question of actual rights (or at least welfare) of those who are being consumed as food might have a much larger following among the members of the Parties for Animals and the Council of All Beings (and some other government and nongovernmental environmental organizations). However, the proxy representation, either species-specific or focused on individuals or ecosystems, will create an actual majority of nonhuman voices. Some trickier questions, such as veganism and natural predation, might arise (e.g., when one proxy represents a lion and another an antelope). There is also a question if deliberation is needed for nonhuman-nonhuman relations, rather than equitably considering the claims of nonhumans in human decisions that will affect them. Some proxies would be addressing only the latter while allowing nature to "do its thing." One could see veganism as especially tricky to the extent that it will be contextual, while the question of meddling in natural predation may be easier to resolve as it comes from a position of human sovereignty over animals [60]. Plant-based and laboratory-grown alternatives are an additional path to resolving conflicts, making voices of "food" heard, and facilitating the context in which those voices could be heard, contesting traditional meat, fish, and fowl industries.

4 Ecocentric Forestry

Similar concerns about monocultures in food production also apply in forestry. In the Netherlands, for example, selective cutting of forest is widespread [61, 62]. Due to Dutch population density and agriculture's pervasiveness, forested areas are often thinned [63]. While forest thinning is not as harmful as clear-cutting, its consequences often go beyond this single action as removing old trees prevents hollows where other animals can nest [64].

Evans and Clark [65] reflect that if Leopold's land ethical implications evolve over the forestry knowledge, foresters would *"face choices that require them to balance the integrity, stability, and beauty of the forest with financial expectations or other human-centered objectives."* The question of expectations is important in connection to proxy representation, as it suggests that multispecies interests can go along with other more conventional objectives. Evans and Clark continue that *"good forestry succeeds when both outcomes are met"* (p. 54). Ecological forestry [61] and agroforestry [66], for example, can create so-called carbon sinks that absorb greenhouse gases, benefitting both biodiversity and indigenous communities [51, 66, 67]. Traditional agroforestry includes small-scale slash-and-burn (fire-fallow cultivation) and swidden agriculture [68]. Within protected nature areas, small-scale cocoa growers allow other plants to remain [69], an example of a "middle" landscape between natural and cultural landscapes associated with positive outcomes for biodiversity conservation.

However, these techniques are not scalable or sustainable for large human populations. Note that tropical forest soils' nutrient-holding capacity is limited. When old growth is cut, the thin layer of fertile top-soil easily erodes, opening the bare surface to fires [68]. This points to the need to look beyond even "traditional" forest use [55, 68, 70, 71].

Another example of complexities involves coppicing, "*a way to produce trees that have multiple stems, rather than a single tree trunk. It is an old tradition of woodland management used for conservation and to produce timber without the need to plant new trees*" (https://letsgrowwild.uk/what-is-coppicing-how-to-coppice-for-firewood/). The issue that comes to the fore is how knowledge of history or technical expertise can aid ecocentric decisions. Considering a theory that people have killed off the wild ox in the past, a mega-herbivore might have kept vegetation less dense [72]. Today, coppicing creates periodic openings in the woodland, attracting individual native butterflies https://letsgrowwild.uk/what-is-coppicing-how-to-coppice-for-firewood/). In this case, coppicing can be seen as an ethical good because even though it is working against the "new wild" state of the uniform closed-canopy forest, it favors native biodiversity. Other experts, however, have argued that Pleistocene extinctions that have caused the dawn of wild herbivores are more challenging to establish in archaeological records than a more recent influence of pastoral communities with their herds of cattle [73–76]. The initial reliance on paleo-ecological data indicating open savannahs has been disputed, arguing that dense woods and not open landscapes might have dominated the landscape before medieval livestock grazing [73, 75]. In this sense, open spaces appear as "cultural landscapes" [73, 74].

Yet, from an ecodemocratic standpoint of mutual flourishing, one could also argue that when there are practices such as traditional woodsmanship (which is opposed to industrial forestry) that shape ecosystem but do so in a way that supports biodiversity (as arguably in the case of woodland coppicing), human intervention is justified. Describing some examples of so-called co-created habitats, Gray [77] investigated lowland meadows, heathland, coppiced woodland, and old orchards, proposing that such places are very different from the large-scale, top-down semi-industrial types of environmental management. These types of landscapes, Gray argues, require simultaneous understanding and affection for nonhuman habitats, but also require the stepping back, creating landscapes as a *"roadmap for a revival of forgotten skills in a future culture of simplicity and creativity"* [77, p. 53]. Indeed, the evidence of sustainable multispecies existence in the past is widespread [71].

On the flip-side, one could also argue that vast wilderness as opposed to "cultural landscape" areas is still desirable because: (a) we are not wise enough to know how to manage everything; (b) having everything human-controlled will lead to a homogeneity of evolutionary direction (when heterogeneity is better); (c) untrammeled lands are wondrous lands. However, the binary between untouched wilderness and human-altered landscapes is becoming less apparent due to the expansion of human presence, as damaged wild landscapes border backyard meadows. While "nature" can be discovered in many localities, one needs to be careful realizing that cultural landscaping could be an excuse for cutting greenery to provide biofuel, which some governments have greenwashed as sustainable [63, 78]. Bioenergy and biofuels do not guarantee carbon emissions cuts as they result in burning biomass such as trees or algae, and "fuel forests" erase wilderness and compete with agricultural land [78, 79]. From this point of view, environmental management of the forest could indeed profit from nonhuman representation in recognizing the needs of different forest-dwelling stakeholders.

This representation may have different applications depending on the type. The precautionary principle has been used in forests (and forestry), for example, in cases of sustainable forest management and biodiversity [40, 41].

5 Discussion: What Do Food Production and Ecoforestry Say About Ecodemocracy?

We have considered above a few examples of decision-making when greater-than-human interests come to play in food production and forestry. These examples show that applying ecorepresentation in practice presents some challenges and opportunities in terms of expert knowledge in integrated sciences. For example, there were "original landscapes" influenced by wild herbivores, creating large open spaces, or was the forest cover denser.

Humans, not trees or bushes, make all these considerations. Returning to the question as to how these ecodemocratic standpoints could be applied in this case, the use of proxies [2] or the extended precautionary principle [39] could caution about stress or pain in plants [80, 81]. When the cause of pain in plants cannot be scientifically supported, the use of the precautionary principle would similarly advise against intervention or would raise the bar for ethical justification for any intervention that may potentially harm plants. For this, the human claim would need to be more urgent than otherwise. This could further complicate coppicing in particular and, more generally, using plants as construction material, decorative hedges, or recreational spaces. While at first appearing ecologically benign, even traditional swidden farming demonstrates that any agricultural activity can be incompatible with biodiversity conservation in the context of expanding population and growing material demands [68], despite the *"myth of indigenous stewardship"* [71, 82]. Stewardship can still be seen as anthropocentric, whether the community is indigenous or not.

In ecodemocracy, these forest-dependent communities will also include nonhuman forest inhabitants. One of the difficulties here, however, is that local human stakeholders will need to allow nonhuman representatives to "speak up," which might have been acceptable at the times of traditional animist religious worship, but less acceptable in communities that have internalized Western anthropocentrism. The precautionary principle will probably lean toward environmental management when commercial interests are primary, avoiding local opposition but failing to protect nonhumans seen as instrumentally useless. Human proxies originating from indigenous, local, and foreign groups (e.g., Western environmental organizations operating in developing countries) might disagree on what is beneficial for the forest, depending on who or what they represent. Enactment of the Council of All Beings workshop with local communities could help to create understanding and acceptance. Still, the question of the balance of needs and wants of different stakeholders remains. More generally, we are left to ponder: can there be ecocentric forestry?

Similar questions can be asked about food production: Is it democratic to eat animals if they are also seen as stakeholders? The question of natural predation is a tricky one, and the centuries-long discussion about whether humans and particularly the products of their labor, such as factories, roads, and buildings, are part of nature [71, 82] remains open. As to the question of whether human omnivorous diets should be condemned any more than a lion's diet, one could respond that it is a question of choice. Today, most humans can have a rich and varied diet that allows for a considerable reduction of meat and dairy products, while a lion would have nothing instead of meat. Another key distinction between a lion's and a human diet is, one could argue, a question of scale.

Simply put, while there are a few thousand lions left in the wild and almost eight billion omnivorous humans. Food production involves industrial-scale farming, slaughterhouses, packaging, food waste, animal suffering and death; high physical and psychosocial risks for slaughterhouse and factory workers and their communities; pollution of water, soil, and air; and land waste [83]. Such suggest that the way one species turns all other species into "products" is by no means ecologically democratic.

Animal representation does not necessarily imply that veganism is the best choice if the animals could vote. It would problematize natural predation (in moral terms, a lion is no worse than an antelope). Many contextual vegans [58] argue for an end to human oppression and dominance over animals, advocating for respect in their use, which comes through minimizing it to the extent possible given context and providing robust rationales for this. When addressing indistinction and not removing ourselves completely from predation, one could suppose that it creates a false comfort that betrays veganism since living inevitably entails consuming life and problematizing that shared zone and our rationale for engaging in it. In this respect, a contextual rather than a complete universal approach to vegetarianism may be more appropriate in ecodemocracy [33]. In that light, the lion cannot minimize its use of an antelope, but many humans can always keep striving to reduce their use of animals.

Companion animals such as dogs and cats who consume large amounts of meat or even the fish industry that uses animal protein to feed their fish are far more problematic. On the other hand, one could argue that a vegan diet is not likely to be democratically chosen by most existing human electorates or, hypothetically, nonhuman predators. One could say that many nonhuman predators are contextual vegans. For one, animals could vote that humans should be vegan but not the nonhuman predators because humans have a choice while lions and wolves hardly do, and because most of the anthropogenic damage far outstrips consumption by wild predators. From a universalist conception of veganism rather than a contextual one, there is the possibility that veganism is not the best or a viable solution. However, veganism also says something about the necessity of basic rights and particular care for specific groups (ethnic minorities or wild predators). Also, it raises the question of who is represented (cows, pigs, chicken, farmed fish, krill, forests cut down to produce soy) and how much weight they have in the democratic process. Another argument is that human and nonhuman voters could not simply force human individuals to become vegan outside of an authoritarian state. They could still influence the legal foundations and state subsidies that make mass meat production possible and profitable. It raises a significant question about our current democracy, which is about human sovereignty over animals [60], rather than how we can shape society together. With such a complexity, ecodemocracy does imply that food needs to be in direct contact with ethical and pragmatic concerns, all the more apparent at the time of a zoonotic disease outbreak COVID-19. In this sense, ecodemocracy will serve much more than the need to change diet. As Gray has stated, *"our agricultural practices must be re-shaped to support and mesh harmoniously with nonhuman life, rather than obliterating it"* [77, p. 53].

Of course, this does not mean that some forms of the ecodictatorial elite will dictate to people what to eat, how to use wood, or how to live their lives. Bringing animals' and nature's voices into the deliberation processes does not imply overruling humans or forcing them to do certain things but finding a means to develop living conditions to facilitate ecological and ethical lifestyles.

However, it does imply that several assumptions and values need to be re-examined, not in a simplistic win–win scenario of convergence theory, but in a critical and discerning way.

6 Ways Forward

As values have changed and new institutions have indeed emerged, there is room for hope that secures human rights, abolishes slavery, protects individuals, and groups against racism, sexism, and other forms of discrimination. The enduring process of correcting civil and human rights abuses is ongoing. The institutional mechanisms, as well as grassroots rights movements coupled with gradual but progressive actions by parliaments, national courts, and supranational organizations, are not perfect. Still, they do exist in democratic systems without the use of force or dictatorship. The proxies' representation may need to be about minority rights, proportionate to the number of individuals within the species and/or natural requirements for each species (e.g., the amount of territory a predator needs for hunting) bio-proportionality [84]. It might lead to a truly representative democracy that encompasses our human requirements and billions of others' needs.

Yet, before we turn to practicalities such as the question of forestry versus forest or natural predation versus human veganism, we need to address the larger requirements of ecodemocracy. This requirement recognizes the grave injustice that billions of living beings have no "voice" in democratic decision making. This recognition exposes current democracy as vastly inadequate in curbing the planet-wide subordination of places and beings to a single dominant species. To quote Crist: "*Human supremacy has ensconced widespread indifference toward the plight of nonhumans and their homes; it ignores and keeps itself ignorant of, the question of their reproductive rights, as individuals and as species. The dominant culture thus seems unable to grasp the moral evil of erasing wild nature just to accommodate more and more people to live, all at once, on a planet occupied as a resource satellite"* [18, p. 148].

This realization of moral evil translates into the need for urgent action to *"reduce overconsumption and escalating demands"* [50, p. 34]. While the differences between animal ethics and environmental ethics have been exaggerated to mask human exceptionalism and economy-centered and expansionist ideology [34]; an exploration of the relationship between sustainability and the protection of nature and individual nonhumans provides insights into the important role that ethics can play in politics [34]. Once this realization is established, the interests of various stakeholders that currently comprise commodified categories of "food" and "forestry" can be discussed. The Parliament of Things and the Council of All Beings might be good starting points for human beings' underlying ethical positioning toward nature. Adding political bite to The Parliament of Things and the Council of All Beings, Parties for Animals can push the issues forward. However, as their reach is limited and they need to continually compete with existing parties, even a more progressive step toward affirmative action to represent the voiceless is required.

Thus, emanating from the past's civil liberation movements, another key to progress in foregrounding ecodemocracy is establishing institutions that consider the interests (at least the most basic of life) of all other planetary citizens. Existing democratic mechanisms, such as the precautionary principle or Parties for Animals, do not represent all species involved in or affected by food production or forestry. The "precautionary principle" currently exercised is hardly ecodemocratic, as it usually seems to be used technocentrically to avoid harm to human interests (even if these converge with nonhuman claims). Some safe-guard related to food production and its relation to habitat destruction and biodiversity loss [41, 42] are present within the existing democratic systems, but they are limited. Parties for Animals or the precautionary principle mostly applies to food crop safety and soil toxicity but does not necessarily consider surrounding (wild) areas. In this way, we may argue that the prioritization of domestic species over others occurs.

Another challenge is that if the most ambitious form of proxy representation is established, it will need many more representatives than current Green parties, Parties for Animals, or environmental NGOs currently have to represent billions of nonhuman beings. In this case, current anthropocentric electorates need to be augmented by required quota representatives, similar to the previously discriminated minority groups supported through affirmative action. Considering that anthropocentrically minded electorates do not currently elect nearly enough proxies, continuous affirmative action and a quota representation are necessary. To maximize their election results, political organizations belonging to the transnational group of animal parties should also expand their manifestos beyond the restricted single-issue domain. If we follow Calarco's definition of anthropocentrism [58], as well as a(n) (eco)feminist approach [31, 33], these parties would need to be intersectional and advocate for an end to all forms of oppression (racist, classist, xenophobic, gender, heterosexist, ableist, naturist, or anthropocentric). Some have done it already: the Portuguese PAN and the Spanish PACMA, for instance. Others should follow. Extending the discourse on solutions on other important aspects of society could increase their appeals among the electorate and eventually drag new voters to their primary cause.

The reader might wonder about how realistic this aspiration and actions are. Simply, these actions are necessary if our privileged access to food, forests, or anything living on Earth is truly sustainable, both for our future generations and for billions of other Earth inhabitants. And some initiatives have already turned into action, as witnessed by the established mechanisms described above.

Aside from the four forms of existing mechanisms, the precautionary principle, the Council for All Living Beings, the Parties for Animals, and the proxies, a few recent initiatives representing nature or nonhumans have emerged. These include the Nonhuman Rights Project (https://www.nonhumanrights.org/), the Global Alliance for the Rights of Nature (GARN https://therightsofnature.org/), and Global Ecocentric Network for Implementing Ecodemocracy (GENIE). The Nonhuman Rights Project works to secure fundamental rights for nonhumans through litigation, legislation, and education. GARN is concerned with Earth's protection and restoration focusing on enabling Indigenous Peoples to share their unique concerns and solutions with the global community. GENIE is concerned with advocating and implementing nonhuman nature across a range of decision-making processes [85].

Signs of progress are visible in several different contexts. In some countries, including the USA and Australia, rivers and lakes have been granted rights [71]. In June 2020, the Supreme Court of Justice of Colombia declared the Isla de Salamanca National Park a subject of rights to protect it from rampant deforestation (http://files.harmonywithnatureun.org/uploads/upload953.pdf). The same month, the French Climate Council has tabled a proposal to hold a referendum to make ecocide a crime. Implications of such initiatives can be far-reaching.

7 Conclusion

The political and legal initiatives that support nature and nonhuman entities' rights expand the possibilities of democratic governments, existing and emerging, to support nature's flourishing. This chapter's focus on food production and forestry has delved into the four forms of representation. These include expanding the existing precautionary principle, the grassroots organizations such as the Council of All Beings and The Parliament of Things, the Parties for Animals, and proxy representation. These forms demonstrated limitations as well as possibilities of ecodemocracy 'in action.' Examination of (or nonhuman) representation in food production cases and forestry exposed a range of complex questions and the need to ponder trade-offs in decision making. One of the challenges is how representatives can be elected in existing single-species democracies. Examples of existing mechanisms demonstrate that presently not all species affected by food production or forestry are represented. Democratic representation through proxies appointed to represent particular species or collectives such as habitats is furthest-reaching and most ambitious and needs development. It would be challenging to deliberate over how proxies will be elected and how decisions on behalf of various species can be made. Moreover, some questions remain to be addressed. Should all species can be treated as equal or should species ranking be involved? Should representation be proportionate to the number of members of the species or species' significance in ecological chains, and what would the position of humans be in this hierarchy?

Core Messages

- Anthropocentric motivation is inadequate in protecting species and habitats perceived as functionally useless for human welfare.

- This chapter aims to address this gap by inquiring what the practical implications of ecodemocracy using examples of forestry and food production are.
- This chapter raises practical questions about when ecodemocracy is applied in real-life situations and discusses ways forward.
- Some existing and forms of democratic representation of nonhumans are discussed in this chapter.
- Four forms of representation discussed are the precautionary principle, the grassroots organizations such as Council of All Beings or the Parliament of Things, the Parties for Animals, and still a possible one, the proxies.

Acknowledgements This chapter was supported by Reingard Spannring**:** Austrian Science Foundation (FWF, grant number I 4342, 2019-2022).

References

1. Mathews F (1996) Ecology and democracy. Frank Cass, London
2. Lundmark C (1998) Eco-democracy: a green challenge to democratic theory and practice. Dissertation, Umeå Universitet
3. Dobson A (2010) Democracy and nature: speaking and listening. Polit Stud 58(4):752–768
4. Lidskog R, Elander I (2010) Addressing climate change democratically. Multi-level governance, transnational networks, and governmental structures. Sustain Devel 18(1):32–41
5. Gray J, Curry P (2016) Ecodemocracy: helping wildlife's right to survive. ECOS 37:18–27
6. Rensenbrink J (2016) Ecological politics: for survival and democracy. Lexington Books, New York
7. Baxter B (2005) A theory of ecological justice. Routledge, New York
8. Gray J, Curry P (2020) 'Representation for nature': ecodemocratic decision-making as a practical means of integrating ecological and social justice. In: Kopnina H, Washington H (eds) Conservation: integrating social and ecological justice. Springer, Heidelberg, pp 155–166
9. Washington H, Taylor B, Kopnina H, Cryer P, Piccolo J (2017) Why ecocentrism is the key pathway to sustainability. Ecol Citizen 1(1):32–41
10. Washington H, Chapron G, Kopnina H, Curry P, Gray J, Piccolo J (2018) Foregrounding ecojustice in conservation. Biol Cons 228:367–374
11. Piccolo J, Washington H, Kopnina H, Taylor B (2018) Why conservation scientists should re-embrace their ecocentric roots. Conserv Biol 32(4):959–961
12. World Bank (2020) Biodiversity. https://www.worldbank.org/en/topic/biodiversity. Accessed 10 Aug 2020
13. Adelman S (2018) Sustainable development goals, anthropocentrism, and neoliberalism. In: French D, Kotzé L (eds) Sustainable development goals: law, theory, and implementation. Edward Elgar, Northampton, pp 15–40
14. Kopnina H (2020) Education for the future? Critical evaluation of education for sustainable development goals. J Environ Educ 1:1–12
15. Katz E (1999) A pragmatic reconsideration of anthropocentrism. Environ Ethics 21(4):377–390

16. Buckley RC (2016) Triage approaches send adverse political signals for conservation. Front Ecol Evol 4:39
17. Taylor B, Chapron G, Kopnina H, Orlikowska E, Gray J, Piccolo J (2020) The need for ecocentrism in biodiversity conservation. Conserv Biol. https://conbio.onlinelibrary.wiley.com/doi/epdf/https://doi.org/10.1111/cobi.13541 (in press)
18. Crist E (2012) Abundant earth and the population question. In: Cafaro P, Crist E (eds) Life on the brink: environmentalists confront overpopulation. University of Georgia Press, Athens, p 145
19. Kopnina H (2019) Party for animals: introducing students to the democratic representation of nonhumans. Soc Anim 2019:1–21
20. Kopnina H (2019) If animals could talk: ecological citizenship and the Dutch party for animals. Animal Stud J 8(1):158–189
21. Rydén O, Zizka A, Jagers SC, Lindberg SI Antonelli A (2019) Linking democracy and biodiversity conservation: empirical evidence and research gaps. Ambio 1:1–15
22. Eckersley R (1995) Liberal democracy and the rights of nature: the struggle for inclusion. Environ Polit 4(4):169–198
23. Washington H, Cook J (2011) Climate change denial: heads in the sand. Earthscan, London
24. Wilson EO (2003) The future of life. Vintage Books, New York
25. Gonçalves-Souza D, Verburg PH, Dobrovolski R (2020) Habitat loss, extinction predictability and conservation efforts in the terrestrial ecoregions. Biol Cons 246:1–8
26. Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci 114(30): E6089–E6096
27. Leopold A (1948) A sand county almanac, and sketches here and there. Oxford University Press, New York
28. Chapron G, Epstein Y, López-Bao JV (2019) A rights revolution for nature: introduction of legal rights for nature could protect natural systems from destruction. Science 363 (6434):1392–1393
29. Naess A (1985) Identification as a source of deep ecological attitudes. In: Tobias M (ed) Deep ecology. Avant Books, San Diego, pp 256–270
30. Cullinan C (2003) Wild law: a manifesto for earth justice. Green Books, Totnes
31. Warren KJ (1990) The power and the promise of ecological feminism. Environ Ethics 12 (2):125–146
32. Derrida J (1991) 'Eating well', or the calculation of the subject: an interview with Jacques Derrida. In: Cadava E, Connor P, Nancy J-L (eds) Who comes after the subject? Routledge, London, pp 96–119
33. Bailey C (2007) We are what we eat: feminist vegetarianism and the reproduction of racial identity. Hypatia 22(2):39–59
34. Garner R (2015) Environmental politics, animal rights, and ecological justice. In: Kopnina H, Shoreman-Ouimet E (eds) Sustainability: key issues. Routledge Earthscan, New York, pp 331–346
35. Kopnina H, Cherniak B (2016) Neoliberalism and justice in education for sustainable development: a call for inclusive pluralism. Environ Educ Res 22(6):827–841
36. Clugston RM, Hoyt JA (2005) The earth charter and animals. In: Corcoran PB, Vilela M, Roerink A (eds) The earth charter in action: towards a sustainable world. KIT Publishers, Amsterdam, pp 149–150
37. Rockefeller SC (2007) Earth charter ethics and animals. In: Waldau P, Patton K (eds) A communion of subjects: animals in religion, science, and ethics. Columbia University Press, New York, pp 621–628
38. Lynn WS (2019) Rewilding the covenant of life with compassion: a future for global and sustainability ethics. In: Burdon P, Bosselmann K, Engel K (eds) The crisis in global ethics and the future of global governance: fulfilling the promise of the earth charter debate. Edward Elgar Publishing, Northampton, pp 225–245

39. Eckersley R, Gagnon JP (2014) Representing nature and contemporary democracy. Democratic Theory 1(1):94
40. Myers N (1993) Biodiversity and the precautionary principle. Ambio 1:74–79
41. Newton A, Oldfield S (2012) Forest policy, the precautionary principle, and sustainable forest management. In: Cooney R, Dickson B (eds) Biodiversity and the precautionary principle: risk and uncertainty in conservation and sustainable use. Routledge, London, pp 21–38
42. O'Riordan T (2013) Interpreting the precautionary principle. Routledge, New York
43. Marchant GE (2003) From general policy to the legal rule: aspirations and limitations of the precautionary principle. Environ Health Perspect 111(14):1799–1803
44. UNEP (2020) Corona outbreak highlights need to address threats to ecosystems and wildlife. https://www.unenvironment.org/news-and-stories/story/coronavirus-outbreak-highlights-need-address-threats-ecosystems-and-wildlife
45. Macy J (1983) Despair and personal power in the nuclear age. New Society Publishers, Philadelphia
46. Seed J, Macy J, Fleming P, Naess A (1988) Thinking like a mountain: towards a council of all beings. Heretic Books, London
47. Latour B (2005) Reassembling the social: an introduction to actor-network-theory. Oxford University Press, Oxford
48. Whiteside KH (2013) A representative politics of nature? Bruno Latour on collectives and constitutions. Contemp Polit Theory 12(3):185–205
49. Morini M (2018) 'Animals first!' The rise of animal advocacy parties in the EU: a new party family. Contemp Polit 24(4):418–435
50. Garnett T, Appleby MC, Balmford A, Bateman IJ, Benton TG, Bloomer P, Herrero M (2013) Sustainable intensification in agriculture: premises and policies. Science 341(6141):33–34
51. Shiva V (2000) Stolen harvest: the hijacking of the global food supply. South End Press, Cambridge
52. Morris C (2016) Framing the agricultural use of antibiotics and antimicrobial resistance in UK national newspapers and the farming press. J Rural Stud 45:43–53
53. Bowles N, Alexander S, Hadjikakou M (2019) The livestock sector and planetary boundaries: a 'limits to growth' perspective with dietary implications. Ecol Econ 160:128–136
54. Clark MA, Springmann M, Hill J, Tilman D (2019) Multiple health and environmental impacts of foods. PNAS 116(46):23357–23362
55. Cafaro P, Crist E (2012) Life on the brink: environmentalists confront overpopulation. University of Georgia Press, Athens
56. Crist E, Mora C, Engelman R (2017) The interaction of the human population, food production, and biodiversity protection. Science 356(6335):260–264
57. Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on earth. Proc Natl Acad Sci 115(25):6506–6511
58. Calarco R (2014) Being toward meat: anthropocentrism, indistinction, and veganism. Dialect Anthropol 38(4):415–429
59. O'Neill J (2006) Who speaks for nature? In: Haila Y, Dyke C (eds) How nature speaks: the dynamics of the human ecological condition. Duke University Press Books, Durham, pp 261–278
60. Wadiwel DJ (2015) The war against animals. Leiden, Koninklijke Brill nv
61. Batavia C, Nelson MP (2016) Conceptual ambiguities and practical challenges of ecological forestry: a critical review. J Forest 114:572–581
62. Metro (2017) Amsterdamse Bos Heeft enorm bos hout voor de stad [Amsterdam forest has an enormous forest of wood for the city]. www.metronieuws.nl/nieuws/amsterdam/2017/02/amsterdamse-bos-heeft-enorm-bos-hout-voor-de-stad. Accessed 29 July 2020
63. Kopnina H (2019) Plastic flowers and mowed lawns: the exploration of everyday unsustainability. J Urban Int Res Placemaking Urban Sustain 12(2):131–155

64. Zheng Z, Zhang S, Baskin C, Baskin J, Schaefer D, Yang X, Yang L (2016) Hollows in living trees develop slowly but considerably influence the estimate of forest biomass. Funct Ecol 30 (5):830–838
65. Evans AM, Clark FA (2017) Putting the forest first. J Forest 115(1):54
66. Atran S (1993) Itzá Maya tropical agroforestry. Curr Anthropol 14(5):633–700
67. Shoreman-Ouimet E, Kopnina H (2015) Reconciling ecological and social justice to promote biodiversity conservation. Biol Cons 184:320–326
68. Henley D (2011) Swidden farming as an agent of environmental change: ecological myth and historical reality in Indonesia. Environ Hist 17:525–554
69. Schroth G, Harvey CA (2007) Biodiversity conservation in cocoa production landscapes: an overview. Biodivers Conserv 16(8):2237–2244
70. Kopnina H (2017) Commodification of natural resources and forest ecosystem services: examining implications for forest protection. Environ Conserv 44(1):24–33
71. Strang V (2020) The rights of the river: water, culture, and ecological justice. In: Kopnina H, Washington H (eds) Conservation: integrating social and ecological justice. Springer, Cham, pp 105–121
72. Owen-Smith N (1987) Pleistocene extinctions: the pivotal role of mega-herbivores. Paleobiology 13(3):351–362
73. Kirby KJ, Buckland PC, Bullock JM, Hodder KH (2005) Large herbivores in the wildwood and modern naturalistic grazing systems. English Nature Research Reports ENRR648
74. Mitchell FJG (2005) How open were European primeval forests? Hypothesis testing using palaeoecological data. J Ecol 93:168–177
75. Whitehouse NJ, Smith D (2010) How fragmented was the British Holocene wildwood? Perspectives on the '"Vera"' grazing debate from the fossil beetle record. Quatern Sci Rev 29 (2010):539–553
76. Merckx T (2016) Rewilding: an enrichment to Flemish nature conservation. Natuur focus 15 (1):28–33
77. Gray J (2019) Making hay: a conditional defense on ecocentric grounds of various co-created habitats. Ecol Citizen 3(Suppl):43–54
78. Steer A, Hanson C (2015) Biofuels are not a green alternative to fossil fuels. In: The Guardian, January 29. http://www.theguardian.com/environment/2015/jan/29/biofuels-are-not-the-green-alternative-to-fossil-fuels-they-are-sold-as. Accessed 28 June 2020
79. Walsh B (2014) Even advanced biofuels may not be so green. Time, April 21. http://time.com/70110/biofuels-advanced-environment-energy/. Accessed 4 Aug 2020
80. Marder M (2013) Is it ethical to eat plants? Parallax 19(1):29–37
81. Kallhoff A (2014) Plants in ethics: why flourishing deserves moral respect. Environ Values 23 (6):685–700
82. Fennell DA (2008) Ecotourism and the myth of indigenous stewardship. J Sustain Tour 16 (2):129–149
83. Nibert D (1992) Animal rights/human rights: entanglements of oppression and liberation. Rowman and Littlefield, Washington DC
84. Mathews F (2016) From biodiversity-based conservation to an ethic of bio-proportionality. Biol Cons 200:140–148
85. Gray J, Wienhues A, Kopnina H, DeMoss J (2020) Ecodemocracy: operationalizing ecocentrism through political representation for nonhumans. Ecol Citizen 3(20):166–177

Helen Kopnina (Ph.D., Cambridge University, 2002) is currently employed at The Hague University of Applied Sciences, coordinating the Sustainable Business program and conducting research within three main areas: environmental sustainability, environmental education, and biological conservation. She is the (co)author of over 150 peer-reviewed articles and sixteen books.

Francisco J. Santiago-Ávila (Ph.D. University of Wisconsin-Madison, 2019) is a research scientist and instructor at the University of Wisconsin-Madison. Fran’s research is on the integration and application of nature ethics to coexistence with nonhumans, from large carnivores to feral’ species such as cats and horses. As part of this effort, his quantitative research has focused on evaluating the effectiveness of policies and interventions to prevent conflicts with gray wolves, including any policy effects on wolf mortality.

Semantic Analyses of Hilary Putnam's *Twin Earth* (Ft. TE Diagram)

24

C. P. Hertogh

> *Cut the pie any way you like, "meanings" just ain't in the* head!
>
> [2, p. 704]

Summary

Hilary Putnam is one of the most productive thought experimenters of contemporary analytical philosophy. Next to *Brain in a Vat* as derived from some passages in Wittgenstein's *Philosophical Investigations*, Putnam introduced *Twin Earth*, an attack on semantic internalism. On surface analysis, *Twin Earth* appears a sequence of over ten sub-thought experiments, including *Doppelgänger, (Martian) Spaceship, Time Travel, H_2O/XYZ, Aluminum/ Molybdenum*. It is less complex than one might expect, for Putnam proposes an example of a nonmodal analog, too—conceptual confusion of *Elm/Beech*—on deep analysis modus tollens with implicit major, falsifying semantic internalism of *meaning is in the head*. In between the lines, Putnam explains his use of the somewhat disputed modal sci-fi version of *Twin Earth* as motivated by an explanatory gap. One does not need to agree on scientific realism to conclude that philosophers of science should acknowledge the impact of physicochemical (microstructure) theories as 'water = H_2O' and update their examples accordingly. To see to it that (applications of) theories of natural sciences will not

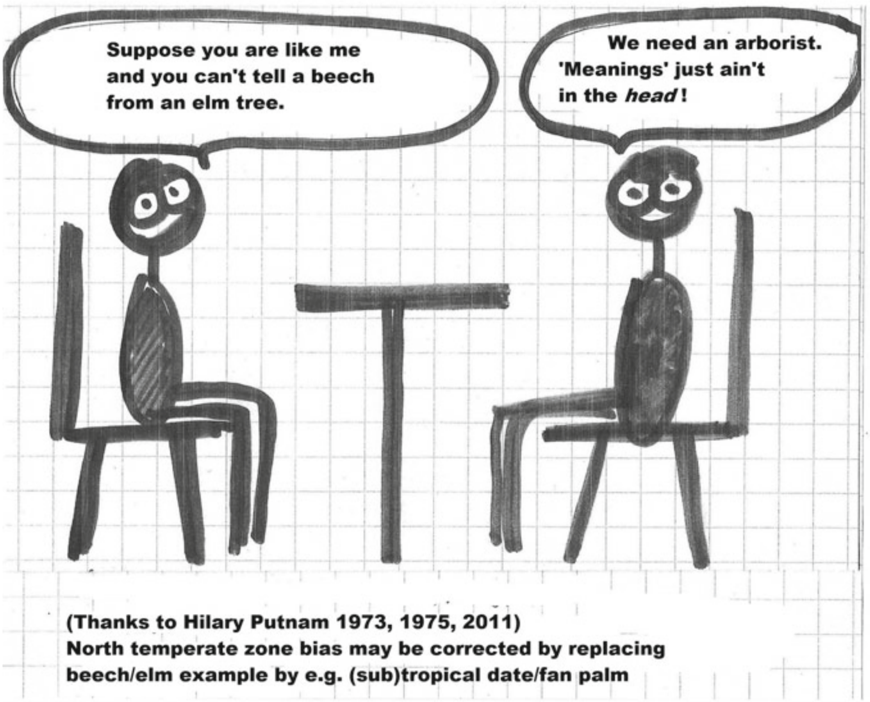
C. P. Hertogh (✉)
School of Foreign Languages and Cultures, Chongqing University (重庆大学), Chongqing, Chongqing, China

Department of Philosophy and Moral Sciences, Free University Brussels (VUB), Brussels, Belgium

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Tehran, Iran

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_24

damage man and nature, there is advanced an update of Karl Popper's progress of science theory unto progress of science and society view by including principles as environmental pragmaticism and global cross-culturalism in the process of error elimination. Two thought experiment methodologies are introduced: TE (Thought Experiment) Matrix, a bracketed logical formula, and TE Diagram, which looks like a Cartesian coordinate system that assigns relative plausibility values to a set of thought experiments. The last sections of the paper explain the methodology of TE Diagram.



'Meanings' just ain't in the *head!* (Putnam 1973, 1975, 2011—semantic externalism)

Adapted with permission from the Association of Science and Art (ASA), USERN; Artist: Maryam Jafari.

The code of this chapter is 01010100 01000101.

1 Introduction

In 'Meaning and Reference' (1973) and 'The Meaning of "Meaning"' (1975), Hilary Putnam introduces a *Twin Earth* thought experiment (TE)[1] to show that meanings are not in the head, as part of a theory called semantic externalism. In

[1] TE: thought experiment(s)
TE_1, TE_2: sub-thought experiment(s)
MTE: mathematical thought experiment(s)
TE_{EI}: Experience TE
TE_{EX}: Experiment TE
TE text: original text of thought experiment (e.g. Putnam [2, 3])

fact, in the section 'Are Meanings in the Head', Putnam introduces both a sci-fi (science-fiction) TE and a non-sci-fi TE ("Suppose you are like me and cannot tell an elm from a beech tree"), with this possibly preventing his theory could fail in the eyes of TE skeptic colleagues because *Twin Earth* could be controversial because of its sci-fi nature.

In Sect. 2, we analyze the surface structure of the TE text of *Twin Earth*, on which analysis there appears a sequence of some ten sub-TE—next to sci-fi *Doppelgänger*, *(Martian) Spaceship*, *Time Travel,* *H_2O/XYZ*, *Aluminum/Molybdenum* TE, a nonmodal *Elm/Beech* TE (Sect. 24.1).

In Sect. 3, logical deep analyses of *Elm/Beech* applying TE Matrix result in modus tollens with implicit major, thereby falsifying semantic internalism by TE minor as 'there is someone who can't tell an elm from a beech'.

In Sect. 4, difference between nonmodal and modal TE is explained as from an explanatory gap or lack of agreement on the exact makeup of real experiments, which may add a generalizing effect to the argument. In Sect. 24.1, the second aim of *Twin Earth* is discussed, scientific realism of microstructure definitions involving chemical formulas as 'water = H_2O'.

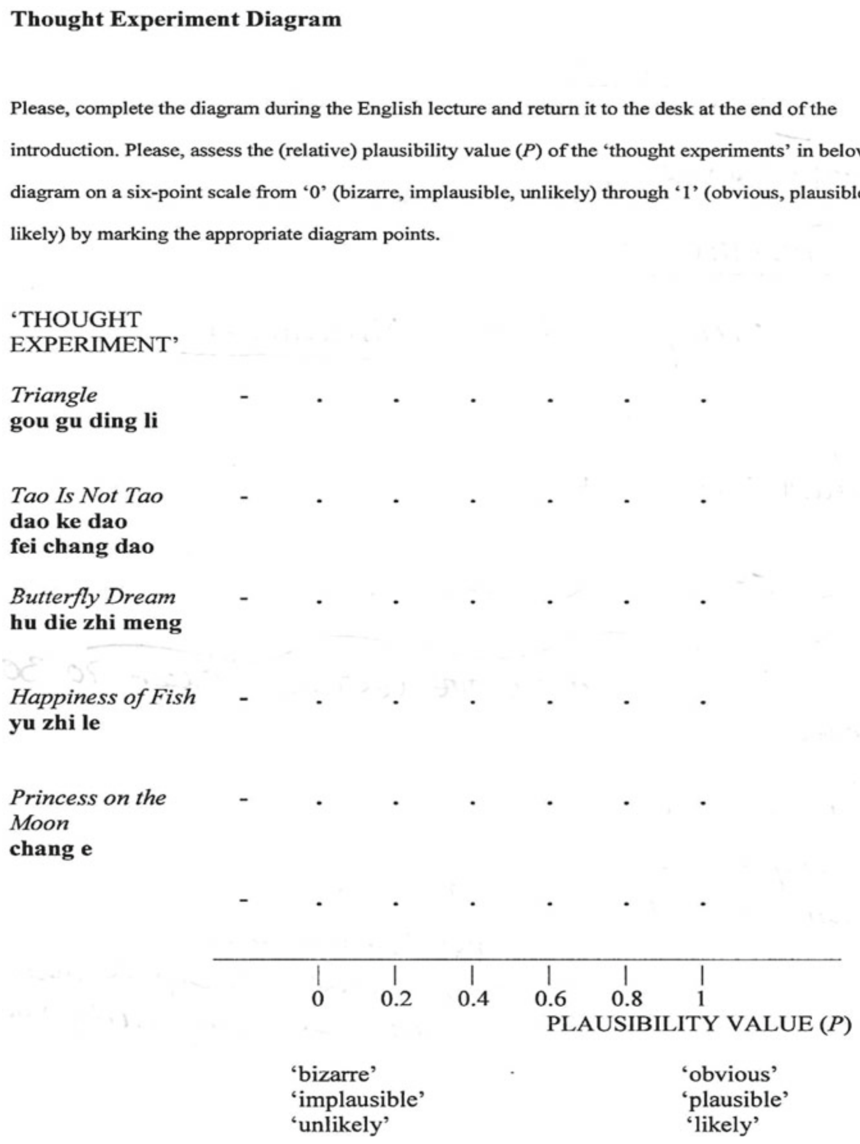
In Sect. 5, it is concluded that disadvantage of microstructure, expert extension definitions as 'water = H_2O' is possible harm to man and nature[2] that may be brought about by the technological application of these theories to society, referring to 1955 *Russell-Einstein Manifesto* about the unforeseen devastating effect of the atomic bomb. In Sect. 24.1, Karl Popper's formula of the progress of science is extended to the formula of the progress of science and society by introducing criticisms from environmental pragmaticism (EP) and global cross-culturalism (GCC) in the phase of error elimination.

In Sect. 6, methodology of TE Diagram survey is discussed, that is a questionnaire that may collect (relative) plausibility (Nicholas Rescher) values from an academic audience with respect to a group of TE. Plausibility values may change from prima face (PF) conceivability to secunda facie (SF) conceivability to ideal (ID) conceivability (David Chalmers). Figure 2 is an example of a TE Diagram as used as part of an interactive presentation on Putnam's *Twin Earth* for XV Philosophers' Rally, Jagiellonian University, Krakow, Poland, in June 2019.

In Sect. 7, the pros and cons of use of qualitative and quantitative methods in philosophy and science are shortly discussed and there is concluded to caution in use of quantitative research in philosophy and social sciences.

TE indicator: e.g. 'Suppose', 'Imagine' etc.
TE Matrix: see e.g. Sects. 3 and 6.
TE Diagram: see e.g. Sects. 6 and 7, Supplements-Figs. 1 and 2
PW(S): possible world (semantics)
P: Plausibility (value)
PF, SF, ID conceivability: prima facie, secunda facie, ideal conceivability.

[2] 'Man and nature' is, of course, as with regard to 'man' meant to embrace both females and males, it is about 'human beings, members of the species Man' (Russell-Einstein Manifesto [4]). In other places in this paper 'her' and 'his' (etc.) are alternated, pluralized etc. Also, Putnam uses both.



Thought Experiment Diagram

Please, complete the diagram during the English lecture and return it to the desk at the end of the introduction. Please, assess the (relative) plausibility value (*P*) of the 'thought experiments' in below diagram on a six-point scale from '0' (bizarre, implausible, unlikely) through '1' (obvious, plausible, likely) by marking the appropriate diagram points.

'THOUGHT EXPERIMENT'

Triangle
gou gu ding li

Tao Is Not Tao
dao ke dao
fei chang dao

Butterfly Dream
hu die zhi meng

Happiness of Fish
yu zhi le

Princess on the Moon
chang e

0 0.2 0.4 0.6 0.8 1

PLAUSIBILITY VALUE (*P*)

'bizarre'
'implausible'
'unlikely'

'obvious'
'plausible'
'likely'

Fig. 1 *Zhuangzi* TE diagram—relative plausibility. This TE diagram was designed for an interactive presentation to Course *Dao De Jing* (Prof. Weixue Li), Huxi Campus, Chongqing University (CQ/CN) on May 15, 2013, 19:00–19:30 (CST). See also, e.g., Table 1 and Sect. 24.2

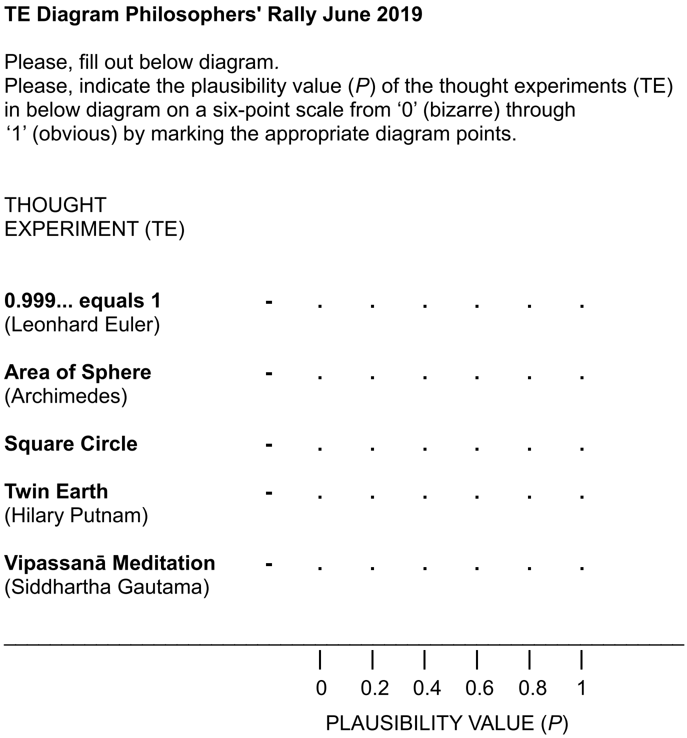


Fig. 2 *Twin Earth* TE diagram—relative plausibility. This TE diagram was designed for interactive presentation to XV Philosophers' Rally, Institute of Philosophy, Jagiellonian University Krakow (PL/EU) on June 29, 2019 (CEST). See also, e.g., Sect. 24.4

2 Surface Analysis

Before discussing the non-sci-fi *Elm/Beech* TE in detail, we will cite the original text of the TE that is introduced 'with the aid of a little science fiction'.

> That psychological state does not determine extension will now be shown with the aid of a little science fiction. For the purpose of the following science-fiction examples, we shall suppose that somewhere in the galaxy there is a planet we shall call Twin Earth. Twin Earth is very much like Earth; in fact, people on Twin Earth even speak *English*. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is *exactly* like Earth [TE_1]. He may even suppose that he has a *Doppelgänger*—an identical copy—on Twin Earth, if he wishes, although my stories will not depend on this [TE_2] [3, p. 223—bracketed counting of sub-TE added].

Twin Earth has some ten sub-TE, which are introduced by TE indicators like 'Suppose', 'Let ('s/us)' i The major sci-fi TE is the supposition of *Twin Earth*, TE_1

ii The second TE, TE_2, seems to be added for dualists among the readers: *if you are a dualist, then also suppose my Doppelgänger* … [3, p. 227]

iii The third TE supposes that *water on Twin Earth is XYZ* instead of H_2O [3, p. 223]

iv The fourth TE imagines a spaceship from Earth visiting Twin Earth as *if a spaceship from Earth ever visits Twin Earth…* [3, p. 223]

v The fifth one is the reverse meant that *symmetrically, if a spaceship from Twin Earth ever visits Earth…* [3, p. 224]

vi The sixth one is a *Time Travel* TE: *Now let us roll the time back to about 1750. At that time chemistry was not developed on either Earth or Twin Earth…* [3, p. 224]

vii It is supported by a *Counterpart* TE involving Oscar_1 and Oscar_2: *Let $Oscar_1$ be such a typical Earthian English speaker, and let $Oscar_2$ be his counterpart on Twin Earth* [3, p. 224]

viii Next, the same complex TE enrolls with regard to aluminum/molybdenum, analogous to H_2O/XYZ: *Let us now modify our science-fiction story. I do not know whether one can make pots and pans out of molybdenum; and if one can make them out of molybdenum, I don't know if they could be distinguished easily from aluminum pots and pans* [3, p. 225]

The *Spaceship* and *Counterpart* TE reiterate, which brings our count to about ten sub-TE.

Meanwhile, Putnam explains why ostensive definitions and 'application of some "common sense" test' [3, p. 225] fail. And Putnam cannot escape from introducing some sort of 'expert knowledge'.

> So, I shall suppose that molybdenum pots and pans *can't* be distinguished from aluminum pots and pans save by an expert [3, p. 225].

One needs to be 'chemically or metallurgically sophisticated' [3, p. 225] to tell the difference between, e.g. H_2O and 'watery stuff' or between aluminum and molybdenum (in the eighth TE).

In the next section of 'The Meaning of "Meaning"' Putnam goes on to explain expert knowledge by '[a] socio-linguistic hypothesis' of '*division of linguistic labor'* [3, p. 227]. Nevertheless, we wonder if the difference between an 'average speaker' or 'standard speaker' and an 'expert' won't amount to something like elite knowledge on part of the latter. It could have been exactly this consideration that may have prevented philosophers like Wittgenstein and Russell from departing from ostensive definitions of natural kinds in an earlier stage of analytical philosophy.

2.1 *Elm/Beech* TE

In 1973 'Meaning and Reference' Putnam also imagines in a non-sci-fi TE someone like him, e.g., the reader, who cannot tell an elm from a beech tree.

> Before discussing this example further, let me introduce a *non*science-fiction example. Suppose you are like me and cannot tell an elm from a beech tree [2, p. 704].

Later Putnam reinforces the argument by imagining

> … then we can always refute him by constructing a 'Twin Earth' example—just let the words 'elm' and 'beech' be switched on Twin Earth (the way 'aluminum' and 'molybdenum' were in the previous example) [2, p. 704].

And Putnam concludes

> Cut the pie any way you like, 'meanings' just ain't in the *head!* [2, p. 704].

Elm/Beech is a three-step TE using TE indicators as 'Suppose ... ', 'constructing a *Twin Earth* example ... ' and again 'suppose ... '. The aim of the TE is the same, showing that meaning is not in the head.

In 1975 'The Meaning of "Meaning"' the TE is identical to the 1973 text. In the end, the non-sci-fi example is introduced and amplified to the sci-fi *Twin Earth* example. In both texts *Twin Earth* TE is an ambitious sci-fi TE that encompasses a *Doppelgänger* TE ('Oscar_1' and 'Oscar_2'), a *Spaceship* TE ('if a spaceship … ever visits …') and a *Time Travel* TE ("Now let's roll the time back to about 1750 …").

3 Deep Analysis—Modus Tollens with Implicit Major (Falsification)

The *Elm/Beech* argument can be reconstructed as modus tollens with the help of higher-order predicate logic, involving relation *Pxy*, 'x has Psychological states (of) y', supposing *x* is a human being and *y* is a psychological state, concept etc. The major states a supposedly universal law of semantic internalism (i.e., that if someone has psychological states, then her/his concept of elm is different from his/her concept of beech), the minor is a counterexample to this supposedly general law (i.e., there is someone who has psychological states and whose concept of elm is identical to his concept of beech), and the conclusion says semantic internalism is false (i.e., it is not true that for all entities with psychological states the concept of elm is different from the concept of beech).

Suppose

Pxy	x has Psychological states (of) y
$\forall$	universal quantifier
$\exists$	existential quantifier
$\rightarrow$	(material) implication
=	equivalence
$\neq$	nonequivalence
–	negation

b	(concept of) beech
e	(concept of) elm.

$$\begin{array}{ll} & \\ (1) & \left[\begin{array}{c} \exists\, xy\, (Pxy \rightarrow (Pxb = Pxe)) \\ \hline -\forall xy\, (Pxy \rightarrow (Pxb \neq Pxe)) \end{array}\right] \end{array} \qquad \langle 1 \rangle$$

(2)

$$\begin{array}{ll} (0') & \forall\, xy\, (Pxy \rightarrow (Pxb \neq Pxe)) \\ (1') & \exists\, xy\, (Pxy \rightarrow (Pxb = Pxe)) \\ \hline (2') & -\forall\, xy\, (Pxy \rightarrow (Pxb \neq Pxe)) \end{array} \qquad \langle 2 \rangle$$

[3]

The first formula, <1> , is a 'raw' TE [6, p. 5], which is still an invalid logical argument, for which reason the formula is bracketed.

Like in Descartes's *Cogito*, the major or universal law, rule etc.—in this case, semantic internalism (see formula <2> , premise (0'))—is hidden–in fact, presupposed or assumed—and the non-sci-fi TE only says 'suppose there is someone whose concept of elm is not different from her concept of beech'—see formulas <1> and <2> , premises (1) and (1') which are identical—from which it then follows that the general law of semantic internalism is false—see formulas <1> , <2> , premises (2) and (2') which are identical, too—and, so, meanings ain't in the head. Formula <2> is like (i.e., is reducible to) modus tollens (including quantifiers), which is a logically valid form of argument. In logic, it is well-known as disproof, proof by contradiction etc., in the philosophy of science as falsification, falsifiability, associated with Karl Popper in twentieth-century western philosophy.

The counterexample can be found on this earth—someone with a different conceptual scheme e.g. someone from Amazonia—as elms are trees that only grow in north temperate zones (please, see *Webster's 11th* definition of elm, which is a near-biological definition)[3].

More easily to imagine for inhabitants of (sub) tropical climate zones is a TE like this

> Suppose you are like me and you can't distinguish a fan palm from a date palm.

---

[3] *Webster's 11th*

> elm …: … any of a genus (Ulmus of the family Ulmaceae, the elm family) of usually large deciduous *north temperate-zone trees* with alternate stipulate leaves and fruit that is a samara …. (italics added).

or

> Suppose there is someone who can't distinguish a fan palm from a date palm.

The distinction is not very difficult to understand since the names of the trees already explain the former has fan-shaped leaves, and the latter dates; like the crucial difference between elms and beeches because the latter have beechnuts and the former not.

Twin Earth could be imagined as just another climate zone of Planet Earth where everything is or seems the same except for age-old civilizations, local cultures and customs, flora and fauna, and the trees are not foremost elm and beech (a/o), but, e.g., fan palm and date palm (a/o).

In the *Twin Earth* TE semantic externalism may not want to do more than explicating a hidden presupposition of semantic internalism, e.g., that is presupposed that we live in the same external world like John Searle [7] has once explicated quite a few presuppositions or background assumptions (such as gravity) of a statement like 'The cat is on the mat'.

These TE are similar to *Qualia TE* since they imagine a counterexample to general psychological laws. In the *Qualia TE* the counterexamples are not just imaginary but possibly neurological cases of, e.g., color blindness (e.g. Chalmers 1996 [8]).

Twin Earth is also akin to Willard Quine's 1960 (Anti) Radical Translation *Gavagai* TE [9, pp. 28–30] as it confronts a situation wherein people with different conceptual schemes meet, e.g., a North American researcher and an inhabitant of Amazonia.[4]

However, the point of Putnam appears to be not just about different conceptual schemes and microstructure extensions as 'water = H_2O', but about scientific knowledge and 'division of linguistic labor' [3], since Putnam contends in 2011 [10] that expert knowledge may also be required for average native speakers to tell an elm from a beech tree.[5]

---

[4] *Gavagai* TE as introduced in Quine's 1960 *Word and Object* (Quine 1960, pp. 28–30 [9]) demonstrates—in Quine's terminology—indeterminacy of translation as related to inscrutability of reference and ontological relativity.

> The indeterminacy of translation has been less generally appreciated than its somewhat protean domestic analogue. In mentalistic philosophy there is the familiar predicament of private worlds. In speculative neurology there is the circumstance that different neural hookups can account for identical verbal behavior. In language learning there is the multiplicity of individual histories capable of issuing in identical verbal behavior. Still one is ready to say of the domestic situation in all positivistic reasonableness that if two speakers match in all dispositions to verbal behavior there is no sense in imagining semantic differences between them. It is ironic that the interlinguistic case is less noticed, for it is just here that the semantic indeterminacy makes clear empirical sense [9, p. 79].

[5] 2013 transcript of Putnam's [63] 2011 Rolf Schock lecture reads

> You may have noticed that I just shifted from 'meaning' to 'reference'; they are not the same, but they are closely connected. Consider: typical modern English speakers have only a vague idea of what an elm tree looks like. They depend on other people, especially

4 Explanatory Gap (Modal and Nonmodal TE)

Quite a few TE are—like Putnam's *Twin Earth*—only well-known in their modal or sci-fi version, e.g., Thomas Nagel's *Bat* [12], Robert Kirk's [13] and David Chalmers's *Zombies* [8], (e.g. Chalmers's) *Qualia TE* [8], Albert Einstein's *Chasing a Beam of Light* [14].

We can perform the nonmodal, non-sci-fi *Elm/Beech* TE in reality. e.g., setting up an informal linguistic experiment, asking a group of people what they think is an elm, what a beech (or what a date palm and what a fan palm) by showing them pictures of elms and beeches (or fan and date palms) and if they—or some of them —confuse them, conclude that there exist people who cannot tell the difference from observable characteristics of the two trees. We do not need a complex experimental setting for we only have to show there are indeed people who confuse trees like elms and beeches or fan and date palms. In fact, this informal experiment has been conducted during an interactive presentation of XV Philosophers' Rally in 2019, and the results were as assumed in Putnam's TE (see Sect. 24.4).

The *Elm/Beech* TE is rather an *Experience TE* (TE_{EI}) than an *Experiment TE* (TE_{EX}) as Putnam may only be appealing to the experience that Putnam himself, or someone else, sometimes confuse referents of similar daily concepts, and Putnam may correctly assume the reader may have the same experience.

However, Putnam adds the modal, sci-fi argument of *Twin Earth* reinforced by a *Doppelgänger* TE.

> …. If someone heroically attempts to maintain that the difference between the extension of 'elm' and the extension of 'beech' in *my* idiolect is explained by a difference in my psychological state, then we can always refute him by constructing a *'Twin Earth' example*. Just let the words 'elm' and 'beech' be switched on Twin Earth (the way 'aluminum' and 'molybdenum' were in the previous example). Moreover, [I] suppose I have a *Doppelgänger* on Twin Earth who is molecule for molecule 'identical' with me [(in the sense in which two neckties can be 'identical')]. If you are a dualist, then also suppose my *Doppelgänger* thinks the same verbalized thoughts I do, has the same sense data, the same dispositions, etc. It is absurd to think *his* psychological state is one bit different from mine: yet he 'means' *beech* when he says 'elm' and *I* 'mean' *elm* when I say elm. Cut the pie any way you like, 'meanings' just ain't in the *head!* [3, pp. 226–227, and 2, p. 704, bracketed word and phrase are not in the latter].

But it is not essential to his argument of refuting the thesis of semantic internalism, since the *Elm/Beech* TE is sufficient to this aim and the TE will recall the same experiences of the reader, which moreover can be reinforced by a real linguistic experiment as outlined, possibly involving two types of palms instead of in the north temperate zone more familiar elm and beech.

---

experts, to determine for them which trees are 'elms', and if there is a an island somewhere in which a dialect of English is spoken in which the word 'elm' is the name of a different species of tree, say of beeches, then we would say that on that island 'elm' means *beech*. For natural kind words, e.g., names of biological species and names of substances, difference in reference *counts as* difference in meaning [63, pp. 197–198].

Lastly, we do not know of philosophical contributions that specifically refute (only) the nonmodal *Elm/Beech* TE, so that refuge to the modal variant seems required. Only on a definition of a (strong) TE as essentially unperformable in reality (as Albert Einstein's *Chasing a Beam of Light* [14] which is physically impossible), we may consider to jump to the modal variant immediately, but commentators who do so, do not, in fact, always do so for this reason; quite often the nonmodal, non-TE version is not even noticed in secondary literature—e.g., nonmodal, non-TE analogs of aforementioned TE are 'visually and hearing impaired people (from birth)', 'other minds' in case of Nagel's *Bat* [12], 'sleep-walkers', 'sleeptalkers' in case of Kirk's *Zombies* [13], 'inverted spectrum' in case of Chalmers's *(Inverted) Qualia* [8].

As there have been some philosophical contributions, particularly criticizing sci-fi TE, we can nevertheless defend sci-fi TE by arguing that these sci-fi TE are used to fill an explanatory or logical gap when there are possibly or apparently too many controversies and discussions on details of scientific makeup to agree on realistic examples. TE philosophers may try and appease possible objections to realistic examples by introducing sci-fi-like TE.

One more rationale to use a modal or sci-fi TE instead of a nonmodal TE is that the former may have generalizing effects. These generalizing effects may already motivate the choice of a TE over a realistic example as according to (e.g.) John Norton [15] and Albert Einstein [14], generalization is one of the benefits of TE. Modal and sci-fi TE share this feature to a logical extreme. However, the generalization of an example or a possible example is quite often not necessary in case of falsification as, e.g., Putnam may, in fact, only need one (1) realist counterexample to refute semantic internalism.

Many TE philosophers have drawn attention to somehow mystifying generalizing effect of TE, which may be explained in many ways, e.g., as side-effect of modality (e.g. [3]), rhetorics (e.g. saying *I*, meaning *we*, *mankind* e.g. Putnam 1973, 1975, 2011 [2, 3, 10], René Descartes e.g. 1637 [16]; 'more art than science', devices of 'persuasion'—Daniel Dennett 1991, 2013a, 2013b [17–19]), induction ('inductive step' from particularity of premises to generality of conclusion—John Norton 1991 [15]), idealization ('highly idealized experimental description'—Ronald Laymon 1991 [20], 'idealized experiments' 'created by thought'—Albert Einstein e.g. Einstein, Infeld 1938 [21]), logical inference (Sören Häggqvist 2009 [22]), synthetic judgments a priori (Immanuel Kant 1781 [23]), conceptual change, social constructivism (Ludwig Wittgenstein 1953 [24], Tamar Gendler 1996 [25]), abstraction, 'seeing laws of nature' (*Platonic TE*—James Brown 1991a, b [26]), narration (e.g. narrative or story aspects e.g. Sorensen 1992 [6], Norton e.g. 1991 [15], Gendler 1996 [25]; e.g. Einstein's CABOL [14]), visualization (e.g. 'picturesque arguments' Norton 1991 [15], geometrical TE e.g. Hertogh 2021 [43], Imre Lakatos 1976 [66]), cognitive tool or instrument (e.g. 'tools for thinking' Dennett 2013a, b [18, 19], 'philosophical figures' CP Hertogh 2015 [27]) argumentation (e.g. counterfactual e.g. Häggqvist 2009 [22]; hypothetical, reductio (ad absurdum) e.g. Nicolas Rescher 1991 [67]; paradox–Roy Sorensen 1992 [6];

'plausibility arguments' David Chalmers 1996 [8]; 'extended argument view' Hertogh 2015, 2021 [27, 43]).

In Putnam's *Twin Earth,* the generalizing effect is attributed to modality, and as it appears in 2011 [10] to *I/we, mankind* rhetorics comparable to Descartes's 'je pense, donc je suis' [16].

4.1 Scientific Realism

Twin Earth TE and semantic externalism are often connected to Putnam's scientism, e.g., scientifically redefining the extension of, e.g., water as 'water = H_2O' instead of an ostensive definition, or enumeration of properties of 'watery stuff', what Putnam calls the 'stereotype' part of meaning, e.g., '*colorless; transparent; tasteless; thirst-quenching;* etc.' [3, p. 269].[6]

However, we cannot consider *Twin Earth* a *Deconstructive TE* as falsification of semantic internalism does not at the same time construct a *particular* scientific theory of 'normal form description' of, e.g., water.[7]

This scientism is only second to the mainline of argument of the TE, which is to argue that

> … the psychological state of the individual speaker does not determine 'what he means' [3, p. 270].

which reduced aim is obvious from the subtitle of the section in which the TE in its 1975 version appears 'Are meanings in the head?'.

The second aim, to know

> … what meaning *is* in any scientifically interesting sense [3, p. 269].

is not established in the strict sense that the (*Spaceship* and *Time Travel*) TE prove a particular scientific theory, viz. contemporary chemistry as 'water = H_2O', to be (transspatially and transhistorically) true or to be truer than any other possible scientific theory of near or far future. However, we can agree on it in the sense that philosophers should update their theories by advances in the sciences like Kripke

---

[6] Please, note, Putnam doesn't mention generally accepted (macroscopic) chemical properties of 'watery stuff', that were already established before—and independent of—microstructure extension definition as boiling point at 100°, freezing point at 0°.

Next, it is a little surprising Putnam doesn't mention the microstructure of aluminum and molybdenum, resp. $_{13}$AI and $_{42}$Mo, possibly because he wants to keep contents of paper as simple as possible, directed at average philosophical readers rather than (tree) biologists (or arborists) or chemical experts. For a chemical account, please, see, e.g., Julia Bursten 2014 [28].

[7] A *Deconstructive TE* is a TE that destructs (falsifies) an old theory and constructs (verifies) a new theory. This kind of TE is discussed in Brown (1991a, b) [26], where they are called *Platonic TE*, through which we may 'see' the laws of nature, but since we are not Platonist (but empiricist together with e.g. John Norton [29, 30, 64]), we have rebaptized these TE as *Deconstructive TE*.

[31] is arguing for with regard to, e.g., ‘water = H_2O’ and ‘gold is the element with the atomic number 79’ [31, e.g. p. 116], which may be summarized like ‘gold = ${}_{79}Au$’.

5 Environmental Pragmaticism (e.g. *Earth Strike*)

One may wonder if these chemical microstructural theories aren’t beyond the scope of philosophy, if philosophy shouldn’t stay with examples from ordinary language, since these theories may involve work of scientists, but have hardly any effect on daily life, as for most people water just remains ‘watery stuff’ and they don’t need knowledge of the chemical constituents to use water in daily life, like the paradigm change from geocentrism to heliocentrism hasn’t effected daily usage of, e.g., ‘the sun rises’ and ‘the sun sets’ (although, of course, nowadays we are aware of different time zones, curvature of the earth etc. e.g. during air travel, long distance communication, using new media technologies of the global village etc.). Next, this issue may embrace political questions as relating to use and abuse of technocratic knowledge by ‘elites’ in society (e.g. ‘an “elite” of bureaucrats and technical intelligentsia’, see Noam Chomsky 1967 [32]).

We may argue that use and application of chemical formulas have become that widespread last decades that it has become part of public education now, which is accessible to everyone. But our doubts are of an environmental nature—although we won’t answer the political questions here, we want to stress that environmental and ecological concerns should be part of the responsibility of foundational scientists and philosophers nowadays, possibly different from last century, that is to say, that philosophers and scientists should strive for progress of science and society, instead of only the former, progress of science, one of the main issues of philosophy of science of last century.

Regarding the example of ‘water = H_2O’, one may fear that the discovery and technological development of the elements of the periodic table, quantum mechanics etc. may have brought about some environmental hazards for Planet Earth, that haven’t been fully coped with, because societal applications of the periodic table as in nuclear energy, chemical industry etc. may damage not only natural environment of the planet, but also human health.

Environmental pragmaticism relates (applications of) progress of science to progress of society and says, e.g., that the more damaging effects (societal applications of) a theory have on human health and ecology of Planet Earth, the less true the theory is (environmentally pragmaticistically speaking). Theory A is less environmentally pragmaticistically true than theory B if (technological) societal applications of theory A cause more damage to man and nature than (technological) societal applications of theory B.

One of the turning points in unlimited progress of science, that has often been damaging human health and Planet Earth’s ecology in recent history—was dropping of an A-bomb on Hiroshima and Nagasaki on August 6 and 9, 1945, by USA air force and its (partly unforeseen) damaging long-term effects on human health and environment.

In 1955, alarmed by (e.g.) incident of dying fish in the nets of Japanese fishermen, Bertrand Russell, Albert Einstein, and some ten more internationally renowned scientists conceived of a manifest, saying

> We are speaking on this occasion, not as members of this or that nation, continent, or creed, but as human beings, members of the species Man, whose continued existence is in doubt… [4].

As a result of the manifest and many conferences of celebrated scientists there came about Nuclear-Weapons-Free Zones, in total, seven treaties, particularly among countries of the Southern Hemisphere.

Some results of scientific discoveries and applications have been banned in recent history (e.g. nuclear arms partly, lobotomy, and more examples wholly), and most of them have been adapted to changing ecological needs of a postmodern society. Some adaptation was already part of the process of science as envisioned by Popper, and in next section we propose environmental pragmaticism may be considered part of error correction.

Environmental pragmaticism is part of progress of science and society, together with global cross-culturalism (as proposed in [27]).[8] Global cross-culturalism relates, e.g., to antidiscrimination, antiracism etc. with regard to accessibility of knowledge to all levels and groups in twenty-first-century societies, that join in the global village. Next, theory A is less globally cross-culturally true than theory B if data from theory A are collected from less variety of cultures than theory B.

These are not political issues yet, but remain within the responsibility of philosophers and scientists, e.g., relating to how they present their theories and views.

Nowadays the environmentally pragmaticist warnings are updated by concerns about climate change because of pollution by chemical gases, notably carbon dioxide (CO_2), from use of fossil fuels etc.

There are quite a few local and global green (grassroots) movements drawing attention to environmental hazards all the time, e.g., environmental activist Greta Thunberg's *Fridays For Future* (FFF) [36], *Earth Strike* ([38] e.g. Noam

[8] Hertogh (2015) [27] discusses identifications, analyses and interpretations of TE in (philosophy of) mathematics, mind, and natural sciences. Global cross-culturalism may particularly relate to humanities, environmental pragmaticism to natural sciences. Update of Popperian formula in Sect. 24.1 appears in Hertogh (2018) on *Vipassanā Meditation*, n. 9 [33, 34] and Hertogh (2020) on Goodman's *Grue*, Sect. 4.7 [35]. In this paper on Putnam it is about EP, Environmental Pragmaticism, in paper on Goodman's *Grue* particularly about GCC, Global Cross-culturalism. Both tendencies have meanwhile been partly materialized in abovementioned local and global movements fighting climate change, and racism e.g. *Black Lives Matter* (BLM)—see e.g. Hertogh (2015), Sect. 1.4 on Martin Luther King Jr. 's 'I Have a Dream' [27, pp. 52–59], involving logic and critical thinking. Two hundred nations agreed on the Glasgow Climate Pact, accelerating climate action and support, during United Nations COP26 conference at the end of 2021 [39], where former USA President Barack Obama had a speech on, e.g., green activism and society [40], quoting Shakespeare's "What wound did ever heal but by degrees" (Othello II, iii, 376–379).

Chomsky), *Extinction Rebellion* [37], fighting climate change by general strikes (e.g. September 2019 climate strikes), school strikes (e.g. on Fridays), civil disobedience, massive protests, sit-ins etc.

Noam Chomsky on Mission Statement of *Earth Strike* mentioned that

> The Mission Statement of the Earth Strike is bold, ambitious, and cogent. More than that, it issues a challenge that is of extreme urgency, addressing the most crucial question humans have ever had to face: do we have the will and the honor to act now to preserve the possibility of a decent existence for future generations, or are we so self-centered and cowardly that we will impose upon them a bitter fate—and not in the distant future [41].

5.1 Growth of Knowledge in Science and Society

Karl Popper is well-known for his work on falsifiability and growth of knowledge. Hypotheses that are developed for testing should use logical formalizations that are (in form at least) both verifiable and falsifiable, and progress of science can't proceed without error elimination.

Popper proposes (e.g.) next formula to summarize his ideas about progress of science

> Using '*P*' for problem, '*TS*' for tentative solutions and '*EE*' for error elimination we can describe the fundamental evolutionary sequence of events as follows:
>
> $P_1 \rightarrow TS \rightarrow EE \rightarrow P_2$ [42, p. 243]

We would like to add ecological and cross-cultural concerns to Popper's formula. If the TS, tentative solutions or theories, are not part of environmental research, the possible negative effects on human health and natural environment should be corrected as part of EE, error elimination.

If we abbreviate environmental pragmaticism as EP and global cross-culturalism as GCC, we may add them to Popper's formula of growth of science like this

$$P_1 \rightarrow TS \rightarrow EE/EP\ GCC \rightarrow P_2$$

Please note that EP and GCC involve effects of TS on society, so, by adding EP and GCC the formula now represents growth of knowledge in science and society since the possibly damaging effects of science on society are corrected in the phase of EE by focusing on EP and GCC.

6 TE Diagram

In Hertogh [27] we have introduced TE terminology, such as TE text, TE indicator etc., and TE methodologies, such as TE Matrix and TE Diagram. In 2015 Glossary, TE Matrix is provisionally defined as a TE typical notation indicating that formulas represent thoughts, propositions only—not facts nor states of affairs in reality. It may be visualized by broad brackets around a formula indicating second feature of TE that it is not a full formally valid argument yet, but still an incomplete, unfinished, elliptical argument, or enthymeme. Application of TE Matrices involves a procedure of some steps, of which the last one is a formally valid (and sound) argument that can do without brackets.

In these analyses of *Twin Earth*, formula <1> (see Sect. 3 of this paper) is the core TE that should be bracketed because it is not a formally valid argument yet (which is, for instance, a reason why some doubt the use and value of TE as a philosophical tool). However, the force of the TE is in its development to a valid argument—above formula <2> without brackets—as by revealing, explicating of hidden or tacit theoretical or contextual premises, presuppositions, principles etc. In the case of *Twin Earth,* it is about hidden major of semantic internalism ((0') in formula <2>), saying if someone has psychological states, then she will know that an elm is not a beech. The (possibly imaginary) counterexample of someone who does not know the difference between an elm and a beech ((1) in formula <1> , (1') in formula <2>) falsifies the hidden major, and, thus, semantic internalism. The argument is next generalized with the help of more examples of natural kinds (as water, aluminum etc.) in the modal *Twin Earth* TE with its many sub-TE.[9] The argument is also understood as highlighting the difference between intension and extension, meaning and reference (cf. title of 1973 TE text [2]).

A second methodology is TE Diagram, which is provisionally defined as a two-dimensional perpendicular diagram, an informal Cartesian coordinate system, involving a vertical line, on which particular TE are set out as discrete coordinates, and a horizontal assessment line of, e.g., plausibility values; the diagram relates particularly TE to relative plausibility values on a many-valued plausibility scale. Plausibility (value), *P*, may be defined as a decimal score between 0 and 1 (e.g. 0.1, 0.2 … 0.8, 0.9, 1) indicating relative plausibility of TE in the TE Diagram. Plausibility value *P* is akin

[9] The possible developments of TE from the wordings of the TE text to its theoretical consequences (quite often at least a little surprising or unexpectedly explanatory in strong TE) by one or more intermediary steps is explained in generic account of MTE (see [43]). In physical sciences, TE usually won't do to prove or disprove a theory or theoretical statement (like a falsifiable hypothesis derived from a theory, e.g., in the hypothetical-deductive method of theory testing), but one could expect real experiments (that can be repeated, double-checked etc. in reality).

The intermediary TE steps may include $[TE]_{RS}$ or TE in restricted sense, $[TE]_{BS}$ or TE in broad sense unto $[TE]_{EX.}$ or extended TE.

TE Matrix is more fully explained in Hertogh (2015), Chap. 16, 'Thought Experiment Matrix and Transformation Rules' [27, pp. 379–386]. TE Diagram is explained in, e.g., Hertogh (2015), Supplement B, of which the account in this paper is a synopsis [27, pp. 416–426].

More theoretical parts of TE theory as involving Chalmers's 2002 prima facie (PF), secunda facie (SF), ideal (ID) conceivability [44] may also be explained in other places, e.g. in the assessment of Einstein's *Chasing a Beam of Light* (which TE is physically impossible) [27, pp. 373–375]. Abbreviations and senses of the terms, such as use of secunda facie conceivability as involving a comparison between TE on a next stage of reflection, are ours, not necessarily Chalmers's.

These considerations are independent of one's view on, e.g., Saul Kripke's redefinition of e.g. a posteriori necessity (e.g. 'Hesperus is Phosphorus', 'Cicero is Tully', 'water is H_2O' [31]), and Chalmers's 2020 theory of two-dimensionalism [45].

but different from probability value *p*–the former (*P*) belongs not just to psychological, but to pragmaticist vocab, the latter (*p*) to mathematical and logical vocab, the former (in a psychological sense) may also be called conceivability value, the latter possibility value (see Hertogh 2015, e.g., Glossary [27]).

TE Diagrams may be used as surveys or questionnaires to get an indication of plausibility of TE within a group, population etc. It may also be used as study tool to test one's individual modal intuitions of TE, e.g., before and after reading of a TE text, checking for oneself if (relative) plausibility values have increased or decreased.

The concept of plausibility value (*P*) as used in the TE Diagrams is derived from Rescher (e.g. [46, 47]) and Chalmers (e.g. [8]).

In a section on *Qualia TE* Chalmers considers TE plausibility arguments

> These arguments from thought experiments are only plausibility arguments, as always, but I think they have considerable force [8, p. 250].

The adjectives 'bizarre' (very implausible) and 'obvious' (very plausible) as used in explanation to TE Diagram are derived from Chalmers as they are used in Chalmers [8, pp. 94–98], where there are offered some considerations on the adjudication of TE, i.e., logical possibility of *Zombies*.

Both Rescher 1976 [46] on plausibilistic inference and Rescher 1978 [47] on Charles S. Peirce's theory of induction define nonanalytical plausibility with reference to Peirce

> By plausibility, I mean the degree to which a theory ought to recommend itself to our belief independently of any kind of evidence other than our instinct urging us to regard it favorably. All the other races of animals certainly have such instincts; why refuse them to mankind? … Physicists certainly today continue largely to be influenced by such plausibilities in selecting which of several hypotheses they will first put to the test'(Peirce in ca. 1910 [48] as quoted by Rescher, partly in [46] and wholly in [47]).

This description comes close to intuition, the mental faculty that is mentioned perhaps most often as an informal explanation of the psychological mechanics of TE.

Two of three senses listed at 'plausible' in *Webster's 11th* have 'superficially', and the third sense 'appearing', which confirm Peirce's and Chalmers's accounts, that it is—in Chalmers's terminology—about prima facie (PF) conceivability [45].

Plausibility is a pragmatist notion as after its etymology, from Latin *plausibilis*, worthy of applause, and an example of the third sense 'appearing worthy of belief' relates it to argumentation, 'the argument was both powerful and plausible'.

It is impartial with regard to any psychological explanation as it only seems to measure the effect. It is a (seemingly) qualitative, pragmatist alternate that may be used next to probable and possible.

A first TE Diagram test may indicate prima facie (PF) conceivability or rather secunda facie (SF) conceivability as the TE Diagram already forces to comparison with more TE. A second test, e.g., after reading and reflection of the TE text and possibly secondary literature, could indicate ideal (ID) conceivability. It may accord with, e.g., Popper's notion of rational discussion as part of (philosophical) problem-solving.

Chalmers summarizes the distinction between prima facie (positive) conceivability and ideal (positive) conceivability with regard to TE

> A typical philosophical thought experiment starts with prima facie positive conceivability. A subject does *not* imagine a situation *in fine detail*: microphysical details are usually left *unspecified*, for example. Instead, a subject imagines a situation with certain important features specified, notes that a situation of this kind appears to verify S, and judges that the remaining details are not crucial: they can in principle be filled in to yield a full coherent conception of a situation that verifies S. For the thought-experiment to yield the intended conclusion, this prima facie judgment must be correct, so that S is ideally positively conceivable. If better reasoning would reveal that the details cannot be coherently filled in, or that the situation does not truly verify S, then the thought experiment will typically fail in its purpose. If the prima facie judgment is not defeatable in this way, however, then the thought experiment succeeds, and S is ideally positively conceivable [44] (from Sect. 2, Positive vs. Negative Conceivability).

Successful TE involves ideal (ID) conceivability, and TE that are only prima facie (PF) conceivable may fail on extended reflection (e.g. comparison with other TE), so they are not secunda facie (SF) conceivable. Next, SF TE may appear defeatable on rational discussion and never reach ideal (ID) conceivability.[10]

Prima facie (PF) inconceivability (cognitive dissonance) can be transformed into ideal (ID) conceivability on rational reflection or discussion of the TE, too, e.g., after (modal) logical analyses and interpretations, whereby a TE premise may satisfy an available and accessible theory. On ideal (ID) conceivability, cognitive dissonance may be relieved and disappear, like in analyses of, e.g., Einstein's *Chasing a Beam of Light* (see n. 10).

In the next (sub)sections, we discuss four *TE Diagram* surveys, Conceivability Test TE *Consciousness* (Sect. 24.1), *Zhuangzi* TE Diagram (Sect. 24.2), *MTE* Diagram (Sect. 24.3), and *Twin Earth* TE Diagram (Sect. 24.4). In Figs. 1 and 2 the reader will find *Zhuangzi* TE Diagram and *Twin Earth* TE Diagram, in Table 1 the scores of the *Zhuangzi* TE Diagram.

6.1 Conceivability Test TE *Consciousness* (EFA-1998)

The very first TE Diagram, called Conceivability Test TE *Consciousness*, was conceived as part of the Amsterdam study group of *European Forum Alpbach* (EFA https://www.alpbach.org/en/ EU/AT) in 1998 and contained TE from David

[10] We may, however, hold that all TE are conceivable, if not on given presuppositions and conditions, then on modified conditions, so that e.g. a *Square Circle* may be conceivable as a *Dynamic Diagram*, wherein its static nature is suspended by addition of dimension of time, in which way it comes very close to *Squaring the Circle* proof. (See Hertogh 2021 [43], and Hertogh 2015 [27], Sect. 3.4, which distinguishes next to TE_{EI} and TE_{EX}, TE_{IM}, TE that are usually considered Impossible, e.g., *Square Circle*, and TE_{OI}, TE that, e.g., may be physically impossible but apply concepts from Objective Imagination, like flying man, e.g., Einstein's *Chasing a Beam of Light*.)

Table 1 Numerical scores of *Zhuangzi* TE Diagram survey conducted in Huxi Campus, Chongqing University on May 15, 2013

TE	Total score per TE (max 116 points)	Average *P*-value per TE (total score/116 rounded up)
Triangle	98.6	0.85
Tao Is Not Tao	76.0	0.65
Butterfly Dream	59.4	0.51
Happiness of Fish	72.2	0.62
Princess on Moon	28.8	0.25

The first numerical column shows a total of all the scores of 116 completed TE Diagrams per TE (individual plausibility (*P*) values per TE ranging from 0, 0.2, 0.4, 0.6, 0.8 to 1, counted together). The second numerical column shows the average *P*-value per TE (total score per TE divided by 116). 71 of 116 TE Diagrams were perfectly completed, fully, and correctly. None of the TE Diagrams had serious ambiguities because of partly (none) or 'incorrect' (45 out of 116) completion of the TE Diagram. The main 'fault' was marking a coordination point on the zero dimension of the vertical line just below the listing of the five TE; these TE Diagrams had six marks instead of five. For a next time, we would omit such an (in fact unnecessary) zero dimension of the vertical line and show exactly 30 (5×6) diagram points (instead of 36) by omitting the bottom row of diagram points. Moreover, the hyphens to the left could be omitted or replaced by a straight vertical since some students had also marked these hyphens instead of diagram points. With a MOE of less than 0.1 (n = 116, MOE = 0.98/sqrt 116 = 0.98/10.77 = 0.09, 9%), statistical results of the survey may be reliable and represent *N*, e.g., the (undergraduate) student population of Chongqing University. *Zhuangzi* TE Diagram may be explained more fully in paper derived from the lecture on *Turn to Butterfly Dream* [1], which was presented on May 15, 2013, together with the TE Diagram survey.
See also, e.g., Fig. 1 and Sect. 24.2
Numerical scores of *Zhuangzi* TE Diagram survey
conducted in Chongqing University (CQ/CN) on May 15, 2013
(N: (undergraduate) students CQU; n: 116 respondents; MOE: 9%)

Chalmers 1996 *The Conscious Mind,* e.g., *Qualia TE*, that was about to be studied in the seminar. However, the seminar did not go on, so the survey could not be conducted [49] and [27, Sup. B1].

6.2 Zhuangzi TE Diagram (CQU-2013—Fig. 1, Table 1)

A second TE Diagram survey was conducted at the end of a presentation on Zhuangzi's *Butterfly Dream* as an intermezzo to a course on *Dao De Jing* by Prof. Weixue Li in Chongqing University (http://english.cqu.edu.cn/ CN/CQ) on May 15, 2013. It took place in one of the large lecture halls of Chongqing University [50].

It is the most successful TE Diagram survey until now because there were over a hundred, in fact, 116 participants, undergraduate students to the course on *Dao De Jing*. The diagrams have been completed very well, more than half, 71, perfectly, fully, and correctly, and the remainder had some minor mistakes in completion, but were still very well legible, and the mistakes did not result in ambiguities of interpretation of the scores of the survey.

Zhuangzi TE Diagram lists five Chinese TE on the vertical line, *Triangle* (*Gougu Theorem* gou gu ding li, 勾股定理) [51, 52], *Tao Is Not Tao* (dao ke dao fei chang dao, 道可道 非常道) [1, 53], *Butterfly Dream* (hu die zhi meng, 胡蝶之梦) [1, 50], *Happiness of Fish* (yu zhi le, 魚之樂) [1, 50], *Princess on the Moon* (chang e, 嫦娥) [1]. On the horizontal line there are six decimal plausibility values, 0, 0.2, 0.4, 0.6, 0.8, 1, from 'bizarre', 'very implausible', 'very unlikely' (0) to 'obvious', 'very plausible', 'very likely' (1) (see Fig. 1, *Zhuangzi* TE Diagram).

From numerical results—see Table 1, Scores *Zhuangzi* TE Diagram—it may be provisionally concluded.

From the two TE derived from Zhuangzi, *Happiness of Fish* (total 72.2, average 0.62) appears more plausible than *Butterfly Dream* (total 59.4, average 0.51). It may be explained from natural sciences as biology that the consciousness of fish is more likely than the consciousness of butterflies since vertebrates are considered evolutionarily more developed than insects.

Curiously, the (according to some interpretations) apparent contradiction *Tao Is Not Tao* scores very high (total 76.0, average 0.65), second after *Gougu Theorem* (total 98.6, average 0.85), a geometrical proof (Chinese variant of *Pythagorean Theorem* $a^2 + b^2 = c^2$, possibly discovered before Pythagoras in Europe), so, it may be expected the naive translation of 'dao ke dao, fei chang dao' is not contradictory but may run like, e.g., 'Tao can be said/named, very saying' (according to, e.g., Google Translate). Since the first lines of *Dao De Jing*, *Tao Is Not Tao* score very high, Daoism may be considered in high esteem (among CQU college students) in China nowadays.[11] That *Gougu Theorem* scores, highest is not

[11] The rise of Daoism is confirmed in articles published in e.g. *The China Quarterly*, *Chinese Sociological Review*, *The New York Times*, indicating that religion including Daoism is on the rise in China e.g.

… China has been emerging as a religious powerhouse to be reckoned with … [54, p. 3].

… Since October 2006 the central government of the People's Republic of China has begun to view Daoism as an important asset that can be mobilized to produce political/ideological benefits and actually has put it into practice immediately after this rhetorical change (n. 7) … [55, p. 61].

n. 7—The ideological change of China's authority is marked by the 'Decisions on Some Key Questions Regarding the Construction of Harmonious Society Made by the CCP's Headquarter' passed by the Sixth Plenum of the Central Committee of the Sixteenth National Conference of the CCP in October 2006. Later, in April 2007, the China Religious Culture Communication Association, that is, the *nongovernmental organization format* of the Bureau of Religious Affairs of the State Council, took the lead to organize the huge International Symposium on Dao De Jing (Tao Te Ching) in Xi'an and Hong Kong, which exemplifies how the Chinese government attempts to promote Daoism for 'boosting the soft power of Chinese culture' [55, p. 62].

Zhuangzi (ca. 369–286 BC) and *Zhuangzi* (ca. 3rd Cent. BC) are the second most important Daoist philosopher and oeuvre after Laozi (ca. 6th–4th Cent. BC), to whom *Dao De Jing* (4th Cent. BC) is attributed.

We attempt an (e.g.) Quinean linguistic interpretation of *Dao De Jing* in presentation on Zhuangzi's TE, 'Turn to Butterfly Dream' for Chongqing University [1].

surprising because maths is next to Chinese, the most important school subject in middle school.

A low score of *Princess on the Moon* (total 28.8, average 0.25) is a little surprising since the mythical story of Chang E, who may live on the moon, is celebrated yearly during Mid-Autumn Festival, although the trustability of the myth is considered very limited as, in fact, no one does believe there is really living a princess on the moon.

The results are consistent with a common academic opinion about trustability, reliability etc., hierarchy of the sciences, formal sciences (maths) lead, next to natural sciences (biology), lastly, social sciences (mythology as part of anthropology).

Most interesting for our Zhuangzi research is to double-check the first result since the main hypothesis of Zhuangzi's TE analyses is e.g.

> Identified by TE mood of dream, we analyze it [*Butterfly Dream* TE] as enthymeme (incomplete argument) that can be completed by hidden analogy revealing main presupposition of animal consciousness. We show similar analyses of Zhuangzi's *Happiness of Fish* [50].

6.3 MTE Diagram (VUB-2015)

A third TE Diagram was conducted during an interactive presentation about non-classical logic for members of the Center of Logic and Philosophy of Science (https://clps.research.vub.be/en/home CLPS), Department of Philosophy, Faculty of Arts and Letters, at VUB Brussels, on Ph.D. Day of January 12, 2015. It took place in a classroom of VUB University and focused on mathematical thought experiments (MTE) [56] and [27, Sup. B2].

Theory and examples of this TE methodology have been explained in the proposal of Ph.D. dissertation for VUB Brussels, and its Supplements contain three TE Diagrams, one to be finished before reading, a second one to be completed after reading, and a third one, aforementioned *MTE* Diagram including methodological explanation [27, Sup. B].

The number of participants (ten to fourteen) to the VUB CLPS Ph.D. Day presentation was far too low for the survey to yield statistically reliable results.[12]

[12] MOE has been calculated for the results of *MTE* Diagram of PhD research presentation for CLPS of VUB Brussels, where from about fourteen visitors to the presentation, ten persons have actively and correctly participated in the *MTE* Diagram survey, completing TE Diagram with six choices, assigning a *P* value to all six MTE (*Euler's Theorem*, *Area of Sphere*, *Square Circle*, *Squaring the Circle*, *Dynamic Diagram*, *Vipassanā Meditation*), marking their sheets with personal identifiers etc. When $n = 10$, MOE = 0.98/sqrt 10 = 0.98/3.16 = 0.31, which amounts to about 30%.

Because of that large a MOE as around 30%, differences as between (high scores as) $P_{0.999\ldots=1}$ and $P_{4\pi r2}$ and (lower scores as) $P_{\text{sq. the circle}}$, $P_{\text{dyn. diagram}}$ and $P_{\text{vipas. med.}}$ are not significant on a statistical interpretation of the *MTE* Diagram as representing the VUB population, Faculty of Arts

6.4 Twin Earth TE Diagram (XV Zlot Filozoficzny-2019—Fig. 2)

A fourth TE Diagram survey was conducted at the end of an interactive presentation on Putnam's *Twin Earth* during conference XV Philosophers' Rally (XV Philosophers' Rally) in June 2019 (http://zlot.obf.edu.pl/ EU/PL). The *Twin Earth* TE Diagram survey took place in a classroom of the Institute of Philosophy, Jagiellonian University, Krakow, Poland, on June 29, 2019 [57].

Next to *Twin Earth* there are listed on the vertical line mathematical thought experiment (MTE), such as *Euler's Theorem* (*0.999 ... = 1*) [58, 43] and *Area of a Sphere* (e.g. Archimedes's proof) [43, 11], because Putnam's main background is logic and mathematics, as well as the philosophical conundrum of *Square Circle* [27, 43], and a global cross-cultural TE, Gautama's *Vipassanā Meditation* [33, 34]. Many more TE Diagrams could have been added, of course, comparing *Twin Earth* to TE from the philosophy of language (e.g. WVO Quine's *Gavagai* [9]), philosophy of mind (e.g. René Descartes's *Cogito* [16], Thomas Nagel's *What It Is Like to Be a Bat* [12], Robert Kirk's and David Chalmers's *Zombies* [8, 13]), philosophy of science (e.g. Albert Einstein's *Chasing a Beam of Light* [14]), more modal TE etc. On the horizontal line of the perpendicular TE Diagram again the same six plausibility values (from 0, bizarre, to 1, obvious) (see Fig. 2, *Twin Earth* TE Diagram).

The interactive lecture, 'Semantics of Hilary Putnam's *Twin Earth*', was announced beforehand in the conference program, and a draft of the presentation had been emailed beforehand to organizers, who had provided a book of abstracts of presentations to the conference.

During the presentation, the TE Diagram survey has been amply introduced and explained.

The TE Diagram consists of only one sheet of paper that is to be completed by respondents by indicating appropriate coordination points, crossings of horizontal and vertical axes, thereby assigning *P* values to the TE (see Fig. 2, *Twin Earth* TE Diagram).

A couple of participants had already answered questions during the lecture presentation as that they also did not know the difference between an elm and a beech (like Putnam asserts of himself in the TE text), nor the difference between cross-cultural analog of difference between date palm and fan palm. The presentation had a slide showing (*Wikipedia* etc.) images of an elm and a beech (next to each other on a slide), and another slide showing a date palm and fan palm, but participants could not tell an elm from a beech, nor a date palm from a fan palm, although the names of elm and beech were translated into Polish (resp. 'wiąz' and 'buk') and *Webster's Collegiate Dictionary* definitions had been added.

There were no objections to the *Twin Earth* TE Diagram, and all seven participants have completed the survey. However—like in the case of the 2015 *MTE*

---

and Philosophy population etc. Sample sizes of about ten (n $\approx$ 10) are too small to yield reliable results; a minimum of forty ($n \geq 40$) is usually required (see [27], Sup. B2).

Diagram—since the number of participants was far less than forty and the margin of error was way too large ($n = 7$, MOE = 0.98/sqrt 7 = 0.37, 37%), no reliable or significant statistical results can be drawn from the survey.

7 Quantitative and Qualitative Methodologies in Philosophy and Science

Disadvantages of TE (or metaphysical 'armchair inquiry' as they are sometimes called) are that they do not live up to nowadays epistemological maxims as repeatability, controllability, reliability, e.g., because of fundamental philosophical problems involving, e.g., solipsism, skepticism about the external world, other minds etc., or doubts about traditional philosophical methodologies as introspection, which, e.g., behaviorists do not consider a reliable methodology.

Disadvantages of the nowadays quantitative methods (as often accompanied by visualized results as charts, graphs, diagrams) are that they *look* more sophisticated, unequivocal, and definite than they *are*. They can easily be abused to deceive, confuse, mislead etc., particularly when explanations about n (sample size, number of participants to survey), N (population size n is supposed to represent), *MOE* (margin of error), data (analysis), survey methodology etc. are missing from presentations of the results. E.g., during the presentations of VUB Doctoral Workshops, ULB-VUB Ph.D. Day Psychology and Educational Sciences, where surveys, graphs etc., were used, explanations of methodology were not seldom missing. Stressing the importance of methodological explanation may remain one of the main tasks of philosophy among the humanities and sciences.

We may provisionally conclude

> quantitative data without qualitative explanations are blind
> qualitative explanations without quantitative data are empty [27, Sup. B2]

We can recommend philosophers not to eschew quantitative methods, graphs, charts, diagrams etc. We can recommend scientists to be careful in presenting quantitative results, graphs etc., and to add an explanatory methodological section to their presentations.

About philosophy classes and conferences, where numbers of students and participants are small, it is quite difficult to find a 'sample' larger than forty persons (which is a minimum number of participants to get significant and statistically reliable results), so that also TE Diagrams may mainly be employed for personal use as to check and double-check modal intuitions as they may change from prima facie (PF) conceivability to secunda facie (SF) conceivability, to ideal (ID) conceivability, e.g., from before to after reading of TE texts.

We could recommend the use of TE Diagrams in (e.g.) undergraduate courses, where the number of students is quite large ($n > 40$), and one may try and draw possible qualitative conclusions from the numerical results in rational discussion.[13]

Qualitative research is considered to embrace, e.g., anthropology, psychology, phenomenology, narratology, case studies etc. Please note TE from themselves already contain visual and narrative elements (so we may apply narratological and semiotic analyses in surface analyses), we discuss psychology and phenomenology of TE in, e.g., Hertogh (2018) on *Vipassanā Meditation* [17, 34, 60, 61], and we would like to add to qualitative research of TE, e.g., hermeneutics (surface analyses), and (modal) logic (classical logic and nonclassical logics as PWS applied in deep analyses), environmental sciences (regarding environmental pragmaticism) and critical thinking (regarding global cross-culturalism).

8 Conclusion

In TE research, it may merit to go back to the original TE texts (ad fontes!), where on surface analyses it may appear that a quite well-known TE as Putnam's *Twin Earth* consists of a series of some ten sub-TE, instead of only one TE. Next, it may appear that many criticisms were already anticipated by the TE philosopher, e.g., that next to possibly disputable sci-fi TE there is also introduced a nonmodal TE, in Putnam's case *Elm/Beech* TE, and that modal sci-fi TE could reinforce nonmodal TE. Sci-fi, modal TE may be used because, e.g., they may have a 'generalizing' effect as TE usually have (cf. e.g. John Norton's 'inductive step' [15, p. 129]) and sometimes because one cannot agree on the exact scientific makeup of a nonmodal analog.

[13] Kantian two-liner (after 'Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe sind blind', *Kritik der reinen Vernunft* [23]) is still provisional, a first effort trying to unite benefits of 'rationalism' and 'empiricism' again, which scientific methods may have got abstracted from each other. TE account is not Kantian, although we discuss e.g. Kant's synthetic judgments a priori in Hertogh (2015), Ch. 4, Sect. 4.3.3 [27, pp. 150–155] and we acknowledge that the very name of 'thought experiment' unites both a 'rational' concept as 'thought' and an 'empiricist' concept as 'experiment', and 'thought experiment' could be considered a contradictio in terminis.

With regard to 'they [visualized results of quantitative methods as charts, graphs, diagrams] can easily be abused to deceive, confuse, mislead etc.'.—fallacious deceptions, confusions etc. may be demonstrated and refuted in academic courses of Critical Thinking, and these and more kinds of bias analyses are nowadays part of scientific methodology.

Margin of error (MOE) is a statistical notion that is used in polling etc. MOE has been explained in Open Lecture for Chongqing University 'Scientific Methodology and Critical Thinking' [65] and has been used in assessment of results of quantitative research (involving questionnaires, data analyses, graphs, PPT presentations) as conducted by freshmen and sophomores of College English 4 courses in the last term of 2012–2013 and the first term of 2013–2014 on topics as A Greener Future, Globalization, Understanding Animals, countries of European Union, USA Melting Pot, Intelligent Vehicles etc [65].

It may be calculated by a formula as MOE = 0.98/sqrt n, where n is the 'sample size', which calculation may be independent from N, the 'population' n is supposed to represent. The sample size is the number of participants to the TE Diagram survey.

On deep analyses, *Twin Earth* appears modus tollens, whereby implicit major of semantic internalism is falsified by Putnam's (imaginary) counterexample in the TE minor—Suppose someone cannot tell an elm from a beech tree. In 2011 it may become evident that it does not need to be an imaginary counterexample when Putnam states, "Typically modern English speakers have only a vague idea of what an elm tree looks like …," since we may need an expert (in this case an arborist) to decide on the difference between elms and beeches (the *referents*) beyond the stereotype *meanings* in the heads of standard speakers. Putnam's hypothesis was confirmed for Polish speakers (and resp. 'wiąz' and 'buk') during an interactive presentation of XV Philosophers' Rally in Krakow in 2019.

Next to the first aim of falsification of semantic internalism, sci-fi, modal versions of *Twin Earth*, *Spaceship,* and *Time Travel* TE are mainly introduced to show transspatial and transhistorical impact of chemical extension definitions, such as molecular formulas based on the periodic table of elements, e.g., 'water = H_2O'. Although the advantage is that philosophers may be urged to update their examples, the disadvantage of these definitions is possible harm to (wo)man and nature as caused by societal technological applications of these foundational theories. There is referred to *Russell-Einstein Manifesto* with regard to unforeseen harm of atomic bomb, and these warnings are nowadays updated by concerns about climate change because of pollution by chemical gases, such as carbon dioxide (CO_2), from use of fossil fuels etc. (cf. diverse governmental, EU, UN etc. policies, Greta Thunberg, *Fridays For Future*, *Extinction Rebellion*, *Earth Strike*, Noam Chomsky a/o).

Karl Popper's formula of progress of science may be updated to progress of science and society view as to allow for criticisms from environmental pragmaticism (e.g. theory A is less environmentally pragmaticistically true than theory B if societal, technological applications of A cause more damage to (wo)man and nature than applications of B) and global cross-culturalism (e.g. theory A is less globally cross-culturally true than theory B if data from A are collected from less variety of global cultures than B) in the process of error elimination, to see to it that theories are not confined to only one culture, and that societal applications of theories will not harm (wo)man and nature.

Next to TE Matrix, i.e., a peculiar type of logical deep analyses, that may show falsification of semantic internalism in case of *Twin Earth* as by modus tollens, TE theory may employ TE Diagram to measure relative plausibility values of a set of TE within a(n) (academic) community.

Examples of the use of TE Diagram surveys show we need to be careful with extending the results to statistical data as quite often the sample size is less than forty participants, so that the surveys will not yield statistically reliable results. Nevertheless, TE Diagrams may be employed for personal use or in small groups to show how modal intuitions may change from prima facie (PF) (in)conceivability to ideal (ID) (in)conceivability (David Chalmers). The TE Diagrams themselves already yield reflection of comparison of (a) particular TE to (a set of) similar TE (secunda facie (SF) conceivability).

Generally speaking, the use of surveys and questionnaires in the social sciences needs to be accompanied by *methodological explanations* containing information about, e.g., *N* (population size), *n* (sample size), *MOE* (margin of error), polling and survey methodologies etc., so one should have caution reading charts, graphs, diagrams etc., results of pollings, surveys, questionnaires etc., and presumed objectivity of statistical data is reduced and compromised by possible methodological constraints and restrictions.

Core Messages

- *Twin Earth*—falsification of semantic internalism: logical analysis *Elm/Beech* sub-TE (modus tollens with implicit major).
- Second aim *Twin Earth*–what meaning *is* in scientific sense: e.g. 'water = H_2O'.
- Technological (ab)use physicochemical theories: environmentally pragmaticist concerns.
- TE Diagram surveys: plausibility assessments of sets of TE by academic communities.[14]
- We encourage quantitative research but stress the import of methodological explanation.

Acknowledgements The author would like to thank:
a. members and visitors of Amsterdam study group of European Forum Alpbach (https://www.alpbach.org/en/ EU/AT), e.g., Prof. Robert de Vries, Psychology, University of Maastricht; Prof. Bob Bermond, Neuropsychology, University of Amsterdam; Prof. Emer. Friso Heyt, Sociology, University of Amsterdam (EU/NL);
b. Chongqing University (http://english.cqu.edu.cn/ CN/CQ), e.g., Prof. Weixue Li (李为学); teacher assistants Xiao Li (肖礼), Kai Li (李凯); Chinese teacher Feng Xiaoling (冯小玲) for help with Pinyin and Chinese characters of *Zhuangzi* TE Diagram; 2012–2013 students and visitors of College English 4 courses and *Dao De Jing* course;
c. VUB Brussels (https://www.vub.be/ EU/BE), 2014–2015 students, members, visitors of Center for Logic and Philosophy of Science (https://clps.research.vub.be/en/home EU/BE);
d. XV Zlot Filozoficzny/XV Philosophers' Rally (http://zlot.obf.edu.pl/ EU/PL), organizers, participants, and visitors of presentation on Hilary Putnam's *Twin Earth* on June 29, 2019, particularly Natalia Zakrzewska (University of Warsaw EU/PL); and
e. lastly, hereby the internet is acknowledged as ongoing inspiration for thinking when 'thinking is like googling' (Marcin Miłkowski 2019 [62], possibly after Newell and Simon2019 [59], XV Philosophers' Rally), and as an inexhaustible resource, where one may be able to find TE texts as Putnam [3, 2] within minutes (internet addresses have been added to some references); in modal logical terminology, the internet is a resource for a (seemingly) infinite set of PW (considered as, e.g., theoretical constructs etc.).

[14] Since numbers of students and academic participants are larger in China ($n > 40$), TE Diagram surveys conducted in China may yield statistically reliable results, e.g., *Zhuangzi* TE Diagram survey conducted after interactive presentation on Zhuangzi's *Butterfly Dream* and *Happiness of Fish* in Chongqing University on May 15, 2013 [1]. Please, see Fig. 1, *Zhuangzi* TE Diagram, Table 1, scores of *Zhuangzi* TE Diagram survey, and Sect. 24.2.

References

1. Hertogh CP (2013b) Turn to butterfly dream. Presentation delivered to Weixue Li's course on Dao De Jing. Chongqing University, Chongqing, May 15, 2013 [PPT]
2. Putnam HW (1973) Meaning and reference. J Philos 70:699–711. http://home.sandiego.edu/~baber/metaphysics/readings/Putnam.MeaningAndReference.pdf. Accessed 5 Feb 2022
3. Putnam HW (1975) The meaning of 'meaning'. In: Putnam HW (1975) Mind, language and reality. Philosophical Papers, vol 2. Cambridge University Press, Cambridge, pp 215–271. Also in: Gunderson K (1975) Language, mind, and knowledge (Minnesota studies in the philosophy of science, vol 7), pp 131–193 https://conservancy.umn.edu/handle/11299/185225 . Accessed 5 Feb 2022
4. Russell B, Einstein A et al (1955) Russell–Einstein Manifesto. Read by Bertrand Russell on July 9, 1955, on press conference in Caxton Hall, London
5. Popper KR (2002/1959) The logic of scientific discovery. First English trans. Routledge, London
6. Sorensen RA (1992) Thought experiments. Oxford University Press, Oxford. http://course.sdu.edu.cn/G2S/eWebEditor/uploadfile/20121224164821178.pdf. Accessed 5 Feb 2022
7. Searle JR (1978) Literal meaning. Erkenntnis 13:207–224
8. Chalmers DJ (1996) The conscious mind, in search of a fundamental theory. Oxford University Press, Oxford
9. Quine WVO (1960) Word and object. MIT, Cambridge
10. Putnam HW (2011) Putnam on meaning and externalism. Acceptance speech Rolf Schock Prize in Logic and Philosophy, Royal Swedish Academy of Sciences, Stockholm, 21 Nov 2011 [Video]
11. Heath TL (ed) (1897) The works of Archimedes. (Orig. e.g. On the sphere and cylinder ca. 225 BC) MacMillan, New York
12. Nagel T (1974) What it is like to be a bat. Philos Rev 83(4):435–450
13. Kirk R (1974a) Sentience and behavior. Mind 83:43–60; Kirk R (1974b) Zombies v. materialists. Proc Aristotelian Soc 48(Supp.):135–152
14. Einstein's *Chasing a Beam of Light* (*CABOL*). Einstein A (1969/1949) Autobiographisches/Autobiographical notes (Trans. PA Schilpp) In: Schilpp PA (ed)(1969) Albert Einstein—Philosopher scientist (3rd ed, first pub. 1949) Open Court/Cambridge University Press, La Salle (IL)/Cambridge, pp 52–53, 49–50; Einstein A (1956) Autobiographische Skizze. In: Seelig C (ed) (1956) Helle Zeit—Dunkle Zeit, In memoriam Albert Einstein. Europa Verlag, Zurich, p 10; Wertheimer M (1959) Productive thinking. Harper and Bros, New York, pp 214–215
15. Norton JD (1991) Thought experiments in Einstein's work. In: Horowitz T, Massey G (eds) (1991) Thought experiments in science and philosophy. Rowman and Littlefield, Savage (MD), pp 129–148 http://philsci-archive.pitt.edu/3190/1/thought_experiments_toc.htm. Accessed 5 Feb 2022
16. Descartes R (1637) Discours de la méthode, Pour bien conduire sa raison et chercher la vérité dans les sciences, Plus La Dioptrique, Les Météores, La Géométrie Qui sont des essais de cette méthode. Ian Maire, Leyden
17. Dennett DC (1991) Consciousness explained. Little, Brown, Boston (Penguin Books 1993)
18. Dennett DC (2013a) Intuition pumps and other tools for thinking. WW Norton, New York
19. Dennett DC (2013b) Daniel Dennett: Intuition pumps and other tools for thinking. Interview DIY Review [Video]
20. Laymon R (1991) Thought experiments of Steven Mach and Gouy: Thought experiments as ideal limits and as semantic domains. In: Horowitz T, Massey G (eds) Thought experiments in science and philosophy. Rowman and Littlefield, Savage (MD), pp 167–191
21. Einstein A, Infeld L (1938) The evolution of physics—from early concept to relativity and quanta. Cambridge University Press, Cambridge
22. Häggqvist S (2009) A model for thought experiments. Can J Philos 39(1):55–76

23. Kant I (2004) Kritik der reinen Vernunft (First Pub. 1781 JH Knoch, Riga) Project Gutenberg http://www.gutenberg.org/cache/epub/6342/pg6342.html. (E-book). Accessed 5 Feb 2022
24. Wittgenstein LJJ (1953) Philosophical investigations. Basil Blackwell, Oxford etc.
25. Gendler TS (1996) Imaginary exceptions: On the powers and limits of thought experiment. PhD diss. Harvard University https://www.academia.edu/945481/Imaginary_exceptions_on_the_powers_and_limits_of_thought_experiment. Accessed 5 Feb 2022
26. Brown JR (1991a) The laboratory of the mind, thought experiments in the natural sciences. Routledge, London, New York (Paperback Edition 1993); Brown JR (1991b) Thought experiments: A Platonic account. In: Horowitz T, Massey G (eds) (1991) Thought experiments in science and philosophy. Rowman and Littlefield, Savage (MD), pp 119–128
27. Hertogh CP (2015) Semantics of thought experiments. Prov. PhD diss. VUB Brussels; thesis is not published yet, but it is (partly) available by email request from the author, noromyxo2005@gmail.com
28. Bursten JR (2014) Microstructure without essentialism: A new perspective on chemical classification. Philos Sci 81(4):633–653
29. Norton JD (1996) Are thought experiments just what you thought? Can J Philos 26:333–366
30. Norton JD (2004b) Why thought experiments do not transcend empiricism. In: Hitchcock C (ed) (2004) Contemporary debates in the philosophy of science. Basil Blackwell, Oxford etc., pp 44–66
31. Kripke SA (1980) Naming and necessity. Basil Blackwell, Oxford etc. (Revised and enlarged version of 'Naming and Necessity'. In: Davidson D, Harman G (eds) (1970) Semantics of natural language. Synthese 22(1–2):253–355
32. Chomsky N (1967) The responsibility of intellectuals. The New York review of books 8.3 (February 23, 1967)
33. Bullitt JT (ed) (2013) Sutta Nipata: The Sutta Collection, 1st ed., 2005 Access to Insight (BCBS Edition). From Sri Lanka Tripitaka Project, courtesy of Journal of Buddhist Ethics. (Orig. e.g. 4th Cent. BC) http://www.accesstoinsight.org/tipitaka/kn/snp/index.html. Accessed 5 Feb 2022
34. Hertogh CP (2018) Logical analyses of Vipassanā Meditation. Contemp Buddhism 19 (2):251–274
35. Hertogh CP (2020) Goodman's Grue—Relativized pluralism and paradigmatic thought experiment (Whack). Human Bulletin 3(1):64–103
36. Greta Thunberg's Fridays for Future (2018–2021) https://fridaysforfuture.org/. Accessed 5 Feb 2022
37. Extinction Rebellion (2018–2021) https://extinctionrebellion.uk/, https://rebellion.global/. Accessed 5 Feb 2022
38. Earth Strike (2018–2021) https://www.earth-strike.com/. Accessed 5 Feb 2022
39. COP26 (2021) UN Climate Change Conference UK 2021. https://ukcop26.org/. Accessed 5 Feb 2022
40. Obama BH (2021) Former President of the United States Barack Obama at #COP26 (Full Speech) [Video] https://www.youtube.com/watch?v=69EMd4csZRY. Accessed 5 Feb 2022 https://www.rev.com/blog/transcripts/barack-obama-cop26-climate-speech-transcript. Accessed 5 Feb 2022
41. Chomsky on Earth Strike (2018–2021) https://www.reddit.com/r/socialism/comments/a0e89l/noam_chomsky_on_earth_strikes_mission_statement/. Accessed 5 Feb 2022
42. Popper KR (1979) Objective knowledge: An evolutionary approach, Rev ed, 1st ed., 1972. Clarendon Press, Oxford
43. Hertogh CP (2021) Computation and visualization thought experiments after Lakatos's heuristic guessing method. In: Trepczyński M (ed) (2021) Philosophical approaches to the foundations of logic and mathematics, in honor of Stanisław Krajewski. Brill, Leiden, pp 271–297

44. Chalmers DJ (2002) Does conceivability entail possibility? In: Gendler TS, Hawthorne J (eds) Conceivability and possibility. Clarendon Press/Oxford University Press, Oxford, pp 145–200
45. Chalmers DJ (2010) The Two-dimensional argument against materialism. http://consc.net/papers/2dargument.html. Accessed 5 Feb 2022
46. Rescher N (1976) Plausible reasoning, an introduction to the theory and practice of plausibilistic inference. Van Gorcum, Amsterdam
47. Rescher N (1978) Peirce's philosophy of science, critical studies in his theory of induction and scientific method. University of Notre Dame Press, Notre Dame
48. Hartshorne C, Weiss P (eds) (1934) Collected papers of Charles Sanders Peirce, vol V. Harvard University Press, Cambridge (MA)
49. Hertogh CP (2011) Conceivability test thought experiments consciousness. Unpublished contribution to 1997 Forum of European Culture seminar on David Chalmers (1996)
50. Zhuangzi (庄子) (2006–2020) Zhuangzi. Trans. James Legge 1891. Chinese Text Project. (Orig. Hundred Schools of Thought Period ca. 369–286 BC) http://ctext.org/zhuangzi. Accessed 5 Feb 2022
51. Suanjing Z (周髀算經) (2006–2021) Chinese Tex Project. (Orig. Zhou–Han 50 BC–100 AD) https://ctext.org/zhou-bi-suan-jing. Accessed 5 Feb 2022
52. Hertogh CP (2020) Mathematical thought experiments. Presentation 31st Novembertagung on the history and philosophy of mathematics. Humboldt University, Berlin, November 26–28, 2020
53. Laozi (老子) (2006–2013) Dao De Jing. Trans. Legge J, First Pub. 1861–1872. Chinese Text Project. (Orig. Warring States ca. 475–221 BC) http://ctext.org/dao-de-jing. Accessed 5 Feb 2022
54. Lu Y (2012) Understanding the rise of religion in China. Chinese Sociol Rev 45(2):3–7
55. Yang D-R (2012) New agents and new ethos of Daoism in China today. Chinese Sociol Rev 45(2):48–64
56. Hertogh CP (2015) Semantics of thought experiments ft. nonclassical logic incl. TE Diagram. Interactive presentation on PhD research for VUB Department of Philosophy and Moral Sciences, Center for Logic and Philosophy of Science (CLPS), January 12, 2015 [PPT]
57. Hertogh CP (2019) Semantics of thought experiments ft. Hilary Putnam's *Twin Earth*. Presentation XV Philosophers' Rally, Institute of Philosophy, Jagiellonian University, Krakow, June 30, 2019 [PPT]
58. Euler L (1822) Elements of algebra, Trans. from the French [by J. Hewlett] with the notes of M. Bernoulli and the additions of M de la Grange, 3rd ed. Longman, Hurst, Rees, Orme, London
59. Newell A and HA Simon (2019) Human problem solving. (First Pub. 1972) Echo Point Books and Media, Brattleboro (VT)
60. Sartre JP (1940) L'imaginaire, psychologie phénoménologique de l'imagination. Gallimard, Paris
61. Hertogh CP (1997) Free summary of JP Sartre's L'imaginaire. Trans. unpublished paper assistantship Philosophical Anthropology VU Univ, Amsterdam, p 1983
62. Miłkowski M (2019) Plenary Lecture. XV Philosophers' Rally, Institute of Philosophy, Jagiellonian University, Krakow, June 29, 2019
63. Putnam HW (2013) The development of externalist semantics. Theoria 79:192–203
64. Norton JD (2004a) On thought experiments: Is there more to the argument? PSA Philos Sci 71:1139–1151
65. Hertogh CP (2013a) Scientific methodology and critical thinking. Open lecture presented on April 11, 2013, at Chongqing University, Chongqing [PPT, Video]
66. Lakatos I (1976) Proofs and refutations, the logic of mathematical discovery. (Edited by John Worrall and Elie Zahar) Cambridge University Press, Cambridge

67. Rescher N (1991) Thought experimentation in Presocratic philosophy. In: Horowitz T, Massey G (eds) Thought experiments in science and philosophy. Rowman and Littlefield, Savage (MD), pp 31–41

C. P. Hertogh has been studying Philosophy at both Amsterdam (EU/NL) universities BPhil *VU University* (e.g. assistant to Philosophical Anthropology) *University of Amsterdam* (M.Phil. Thesis 'Identification, Analysis and Interpretation of Metaphorical Indicative Sentences of Subject-Predicate Form'). In 90s CPH has been participating in seminars on consciousness of Amsterdam *European Forum Alpbach*. In 2015 Ph.D. Thesis has been (provisionally) delivered at *VUB Brussels* (EU/BE), Arts and Philosophy, Philosophy and Moral Sciences—'Semantics of Thought Experiments'. incl. 'Addendum Three Global Cross-cultural Thought Experiments In 00s and 10s CPH has been teaching English in Europe, People's Republic of China, Saudi Arabia, which has resulted in some references to Arabic and Chinese philosophy (a/o).**motto** 道可道 **(from Dao De Jing)**

*First line(s) *Dao De Jing* (Google Translate translation—for more discussion—see Hertogh (2013) on Zhuangzi's 'Butterfly Dream' - [1])

道可道 非常道

Tao can be said,
Very saying.

Sport Sciences: An Ideal Field of Play for Integrated Knowledges

25

Dario Dalla Vedova

We are not students of some subject matter, but students of problems. And problems may cut right across the borders of any subject matter or discipline.

Karl Popper

Summary

Sport is an extraordinary challenge for human physical and scientific knowledge. It lies in the complexity and unicity of open living systems, exceptional and unrepeatable performances, uniqueness of each body and individual characteristics, and redundancy of degrees of freedom (DOFs). Further challenges arise from small numbers of top-level athletes and competitions, the difficulty of carrying controlled experiments, non-falsifiable results, experimentally impenetrable areas, search for valid surrogate endpoints, not direct measurability of some human characteristics and consequent conceptualization of multifactorial constructs, variability and unpredictability of environments, selection and verification of useful models, and heterogeneous, spread, and tacit knowledge. Sport is a multidimensional and complex intersection of Physics, Chemistry, and Biology. Science of complexity and networks, integrated Biology, Big Data, models, and modern techniques brought several answers to old questions and opened new challenging frontiers to improve sport sciences. An integrated approach is proving increasingly valuable and necessary.

D. Dalla Vedova (✉)
Institute of Sport Science "Antonio Venerando", CONI—Italian National Olympic Committee, Largo Piero Gabrielli 1, 00197 Rome, Italy
e-mail: dario.dallavedova@coni.it

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Rome, Italy

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_25

Adapted from International Festival of Paintings for Pediatric Patients (IFPPP) Organized by Health and Art (HEART) Group. Artist: Margherita Pierni, Age : 10/12/2005, City : ITALY, Hospital name : Ospedale del Bambino Pietro Barilla – Parma

The code of this chapter is 01010011 01110000 01101111 01110010 01110100.

1 Introduction

Sport represents an extraordinary field of play to challenge human limits as in physical performance as in scientific knowledge. Challenges for modern sport are the complexity of open living systems, exceptional and unrepeatable performances, uniqueness of each body and individual characteristics, redundancy of degrees of freedom in skeletal movement, difficulty in carrying controlled experiments, non-falsifiable results, small numbers of top-level athletes and competitions, presence of experimentally impenetrable areas, search for valid surrogate endpoints, not direct measurability of some human characteristics and consequent conceptualization of multifactorial constructs, presence of variable and unpredictable environments, the selection and verification of useful models, and heterogeneous, spread, and tacit knowledge [1]. When athletes differing in age, anthropometric and functional characteristics, technique, training, and sports equipment obtain almost the same result to hundredths of a second or millimeters, when long races end at photo finish and gaining or losing the podium is a matter of details, what knowledge can find the *causes* of success or defeat? For a long time, there have been various approaches adopted to explain and improve sports performance, and from here comes the first issue. Are we dealing with a *single sports science* or with *multiple sport sciences*? A core of the matter hides under an apparently innocuous linguistic detail. The use of singular or plural has deeply to do with the conception itself of the unity of science and the scientific method [2] and has crucial practical implications. The Anglo-American, Popperian view of science entails a clear hierarchy between the *hard* science, which is experimental, verifiable or falsifiable, and the *soft* sciences regarding humanities, being theoretical, interpretive, historical, and where experiments are not always under control or repeatable. So, science and his method are only one, singular. However, sport is a multidimensional construct at the intersection of Physics, Chemistry, and Biology, where some areas are experimentally impenetrable, chance plays its role, and the presence of circular causality dismantles simple explanations made of cause–effect chains. Although the concept had already been grasped for centuries, studies on complex systems have finally established that in systems with numerous internal links and degrees of freedom (DOFs) or featured by nonlinear dynamics, it is not always possible to use the Newtonian–Cartesian approach. More clearly, the whole is not always the simple sum of parts, somewhere new information is generated and stored, links, patterns and connections are crucial, and some features present at certain levels are not visible elsewhere. Information is not reducible beyond a certain threshold, so the memory acquired by complex living systems at the highest levels is not detectable in their sub-components. It is not only a matter of ignorance but also of principle: no single level can explain all the others [3, 4]. From Poincarè to Einstein and Schrödinger, from Russel to Popper, from Gödel to Turing, Lorentz and Mandelbrot, an extraordinary heritage of the XX century Science is that human knowledge has impassable limits. No general solution exists to Newton's motion laws for more than two bodies; it is not possible to entirely separate an observed phenomenon by his observer. Even formal axiomatic theories have boundaries of provability. Unlike solar eclipses, the weather becomes unpredictable only after a few days, clouds are better conceptualized as fractals than spheres. Organisms and mechanisms are completely different things, even if sometimes we tend to confuse them. A cell is not the simple sum of atoms; an organ is not the sum of cells; a living being is not the sum of organs. The same unit has completely different properties in vivo, in vitro, or in silico. So, any sport outcome is not just a combination of a knee's range of motion, the oxygen uptake, a lactate threshold, or ski wax. The analytic approach is reductionist and mechanistic and works when building a skyscraper, a mobile phone, or sending the man to the moon. In sport and other disciplines, a powerful working strategy is analytic and reductionist, dividing the system Athlete-Equipment-Environment into single parts to study and improve each one singularly. Taxonomies are normally used in various fields to quantify human tasks and performance [5–11]. Single improvements are supposed to be transferable at different levels and additive. It is natural in our way of facing reality and reasoning linearly, and it works. From here comes the powerful concept of so-called *marginal gains*, a transversal strategy is to improve performance from the extraordinary results of Team GB Cycling during the London 2012 Olympic Games. It does not matter how much every variable impacts the result, our understanding of how this happens, and if they are not homogeneous. Indeed—precisely because we do not know *why*, *how,* and *when*—we strive to optimize all of them together. No matter if it is an aerodynamic detail or a comfortable pillow, because the aggregation of several small gains, none of which singularly dominant, can result in the improvement of overall performance [12]. Cycling, sailing, skiing, bobsleigh and luge, rowing, or tennis are sports heavily influenced by technology and, consequently, by the continuous search for every possible small gain. However, even an unchanging sport as swimming, dominated by physical constraints such as the water density and where—have been banned the technological swimsuits—the athlete relies only on his muscles and technique, benefits are coming from many small details in talent selection, training and technical techniques, and nutrition. The process is visible in the slow but continuous improvement during the years of performances in all swimming styles and distances. A powerful strategy is to perform simulations and sensitivity analysis to assess single variables' importance [13–15]. This is the first practical demonstration of the need for many different pieces of knowledge that must be integrated to work, especially in top-level sports. It is like a sort of puzzle where the pieces are placed on different levels, and it is not easy to get the whole picture. However, even if commonly used, the term *integration* has different interpretations in different contexts and is not always properly defined. Sports science usually alludes to multiple approaches: single features are seen and studied from different perspectives and disciplines [16]. Afterward, the parts are put together to produce a global result where only a single variable depending on the chosen metric, such as a time or a distance, is taken to build a ranking and become a champion. The system is open, complex, and the connection between parts and outcome can be speculated within only certain assumptions. A working strategy is to set intermediate and heterogeneous surrogate endpoints: functional qualities, biomarkers, oxygen consumption, skills, forces, power, speed, elasticity, asymmetries, electromyographic profiles, fatigue, and equilibrium are examples of variables thought to be meaningful markers to predict performance. However, the relationships between balance [17], strength [18], or technique [19], and the elite performance are far to be clear and need to be adequately validated [20–23]. A challenging example is the classification process in Paralympic sports: to guarantee fair competitions, athletes are clustered according to their residual functionality, and minimum impairment criteria are established to assign each individual to classes of a single discipline [24]. Remaining within the visible and measurable, although the human body is physically continuous, disability and motor functions are not. An amputation a few centimeters just above or slightly below a joint radically changes its functionality and, consequently, the athlete's performance obtainable. Ethics aside, the scientific controversy surrounding Oscar Pistorius represents the ideal example of the challenges posed by the need to integrate knowledge in complex and multi-hierarchical systems such as the athlete-sports equipment. Pistorius and few other bi-amputee athletes with their sophisticated double carbon fiber artificial legs are a complex and never seen before interaction between Biology, Physiology, Mechanics, and Technology. Are the missing muscles disadvantageous because there is less available propulsion, or rather advantageous because the system's inertia decreases, the athlete produces less lactate, and has springs instead of joints? How to deal with the technology, material, and geometry of prostheses? How do these characteristics apply to different types of track and field competitions? How much time do they earn? Is it reasonable for these athletes to compete with the able-bodied? It has been claimed that it is a kind of new technological doping [25, 26]. It is impossible to conduct controlled experiments or verify/ falsify some statements concerning the fairness of leaving them to compete against other mono-amputated or non-disabled athletes. Over the years, the scientific and technical controversy has been hot. Numerous discussions were opened about the anatomy, physiology, and running biomechanics of animals and humans, laboratory and field studies have been done worldwide. Many scientific papers published, statements, and rebuttals animated the debate. However, due to the problem's particularity and complexity, a final consensus has not been achieved [27]. The feasibility and replicability of controlled experiments, as required by hard science, is another key point. Unique outcomes with the impossibility of replicate a controlled test is a great challenge for any researcher [28]. It is impossible to re-run a race or send backward the time to change a training program after it has been done. Due to its historical nature, complexity, chaotic nature, limitations in time and resources, practical and ethical issues, it is not always possible to find and test all causal chains and, sometimes, understand why an athlete wins or loses [29]. The impossibility of replication sets insuperable limits on the generalizability of any law because there is no single way to conduct an experiment, finding answers, and falsifying hypotheses. Time and resources are limited; every athlete is different, has a unique personal story, and interacts with a continuously variable environment. Tiny differences between top performances, number, and level of competitors, external factors such as injuries, and the finite number of events in a season can shuffle the cards entailing that the strongest athlete or team will not necessarily emerge [20, 30, 31]. Some questions will remain unanswered and paths unexplored. Doping is a scourge of modern sport; cheating threatens fair competition's very roots and meaning. Therefore, anti-doping is a major challenge and implies the integration of different pieces of knowledge. For ethical and legal reasons, protocols and methods used for other drug investigations cannot research the doping effects on performance: a healthy person should not be used as a guinea pig receiving a potentially harmful drug just to do experiments. However, the lack of scientific methods, poorly controlled evidence about how a single drug translates to performance, the choice of adequate surrogate endpoints to evaluate, understanding if and how it works are all problems difficult to overcome [32]. In complex systems, the whole is not just the simple sum of the parts; rather, the study of causation requires its own conceptualization [33]. So, the integration of knowledge is a challenging and promising strategy for future research in sports.

2 Movement in Sports: A Starting Point

Any sport challenges different human abilities and skills, requiring the execution and control of some movements. Even in static disciplines such as Shooting or Archery, every gesture must be performed alternately, creating and breaking a position and a new balance. So, the study of human movement is transversal for all sports. The living bodies are flexible to perform a wide variety of movements. A crucial property of biological tissues is *viscoelasticity*—the ability to feature at the same time viscous and elastic comportment—on which depend stiffness, loads, movements, and, ultimately, the technique [34–36]. It has redundant functional DOFs in muscles, joints, movements, and at the neurophysiological level. Internal dynamics, viscoelastic and soft tissues, wobbling masses, and not uniform density distribution are additional features. Males, females, and different athletes have unique body shapes, size and anthropometric characteristics, muscle, mass, and fat distribution. Care is required when bodily characteristics are generalized between athletes. Usually, highly customized studies and optimizations are needed. Muscles span over multiple joints and never act in isolation, showing length-dependent properties. Impulses and resulting movements are circles of interaction between the nervous system and sensory environment, so no simple correspondence between a motor task and a motor solution exists. Even when performing the same movement repeatedly, the kinematics is not identical [37]. Due to the nonlinear combination of gravitational and apparent accelerations with the kinetic chains' geometry, great analytical difficulties arise in studying fast and multi-joint movements. The relationship between the innervational impulses and consequently evoked movements is extremely complex and not univocal [13, 37]. When studying the kinematics and dynamics of the limbs in the gravitational field, the mechanically complex interaction between different muscles and the non-constancy of the moment of inertia comes up. So, the equations for movement become complex. As already understood by Bernstein in the early twentieth century, *cannot exist an unequivocal relationship between impulses and movements; movements are possible only under conditions of the most accurate and uninterrupted agreement—unforeseen in advance—between the central impulses and the events occurring at the periphery, and are frequently quantitatively less dependent on these central impulses than on the external force field* [37]. Reduction of DOFs, dimensional compression, and compensations are operated by the coordinative structure [38], but even the same final motor solution may be highly individual. An extraordinary principle of Biology and cybernetics is *equifinality* [39]. It states that for open and adaptive systems, there might be the same state derived from different boundary conditions and different paths. So, the unique motor strategies and variability of single athletes may result in very similar outcomes. The systems show a strong path-dependence making it improbable that only a single and generalizable *best way* exists [40]. Human bodies are noisy and not fully determined systems, both these features being crucial in movement analysis and comprehension. The long practice performed by athletes before competing in sports requiring the execution of accurate motor skills such as tennis and golf is not only a warm-up but rather likely a necessary recalibration of a very noisy and high learning-rate sensorimotor network [41]. From the classic deterministic point of view, the movement variability is considered as noise or error. It comes from adaptations in the musculoskeletal system, changes in the external environment, and bio-variance precluding the accurate replication of technical movements and affecting tests [42]. Instead, the modern and integrated view supposes that the variability may contain important information about nonlinear properties of the system [43, 44], the potential risk of injury [45], some underlying learning and training process, latent pathologies and incomplete recoveries, or environmental changes [46]. Singles' kinetic and kinematic parameters describing the technique behave differently than the whole system, so the association of variability with skilled performance is unclear and intriguing [47]. It is worth considering that the top athletes usually feature performances with lesser variability, and males are more regular than females [48–51]. Not all human features are formulable as simple variables. What is a *sports technique*? How to define, analyze, and teach it [52]? Is it unique or different solutions equally effective are possible [53]? Scientists and practitioners analyze athletes' movements in different ways: the former selecting specific, measurable variables, the latter observing the movements of the whole body. This reflects the ancient debate of medicine between organism and body, whether the disease should be seen regarding single organs or living beings as a whole. The selected point of view consequently affects approaches, strategies, and tools to be chosen. More, after years of practice, even skilled athletes fail to reproduce the same movement [43], and this explains why, despite the increasing use of technology, not all questions have been answered. For its nature, history, and culture, for the presence of a boat, rowing is one of the most studied sports. But the pattern of exerted forces as a function of oars' angles is nonlinear and complex. Due to the interplay of anatomical and biomechanical elements and differences between crew members on the same boat, it seems unlikely that only one optimal profile is identifiable for all athletes. The bridge between performance and forces' characteristics of athletes is until under discussion [54]. Top performance entails that all working variables are close to their optimum, conceptualizable as a multidimensional space with a plateau. When everything works properly, the results are weak relationships between parameters, small ranges of values, and flat responses. This shows the challenges of coaching and researching [13] and makes it difficult to provide useful individual advice [55]. The importance of a single variable is shown only in a negative, asymmetric way because it is identified as a *cause* for a problem only when something is not working as it should. A technical detail or a single food cannot be assumed as *the cause* for the victory but maybe the problem *responsible* for a defeat. Different solutions have been proposed to these issues, once again demonstrating the need for integration of knowledge. The first is related to the role of environmental, organismic, and task constraints limiting the number of potential DOFs and producing coordination [56, 57]. Gravity, gender, anthropometry, muscular strength, power, equipment, or cadence influence sports technique. Along with all other natural constraints and sport rules, they limit the possible outcome but do not determine it. Moreover, some characteristics required from a certain performance do not account for all different individuals [43, 54, 58–60]. With the increased availability of technology and, therefore, of the number of variables that can be acquired when multiple features are measured together, there is a growing need for analytical techniques capable of simplifying data without losing information. High-dimensional data is reduced by principal component analysis (PCA), which looks for patterns and clusters. It represents a good meeting point between scientists' analytic approach and the holistic view of coaches [61, 62]. Open issues in this field are the definition of useful properties of the systems, the selection of testing methods [63] with the choice of experimental protocols [64, 65], detailed comprehension of the nature of motor variability, and the information it may provide about a relationship with performance, development of motor skills, recovery, and injury risk [41, 43, 46]. The dynamical systems approach incorporates a wide range of heterogeneous constraints [66], so the research for the athlete-specific optimum still is the *holy grail* of sports movement analysis [67].

3 A Point of Equilibrium

If Sport means movement, a living being entails the continuous search for static or dynamic equilibrium. This is the place where injury prevention [68, 69], motor control, fatigue, sports equipment, and final performance are closely linked together [70–72]. The sense of balance is a highly dynamic phenomenon [73] not dependent on just one organ. It is even more complex than the single five basic senses of hearing, sight, smell, taste, and touch. Systems and organs must cooperate in Mechanics and Biology's synergy to obtain equilibrium. Processing information coming both from the external (exteroception) and internal (proprioception) environments, quickly looking for an adequate solution, they are merged in the *mechanobiology* [74, 75]. Proprioception is decisive for equilibrium, posture, and gait but hardly graspable [76] because it requires a close and complex coordination of different apparatuses. The somatosensory network processes peripheral information coming from heterogeneous subsystems and relative positions of body segments. Muscles, tendons, skin, and joints contain mechanoreceptors able to record pressure, vibration, temperature, and pain. Proprioceptive information is crucial in building the final human balance [77–79]. For about a century, non-continuity of motion control has been stated when the brain converts the sensory data into appropriate motor solutions [37]. In fact, the process requires enough time to be properly performed. The muscular reference system has many dimensions. Possible solutions are not univocal. Simplifying strategies for the neuro computation are required. A powerful strategy is to cut information, making the continuous as discrete. It has been stated that our thoughts run at 40 Hz, but we can act at most about 10 Hz [80], meaning that about 25 ms are necessary to process brain data and 100 ms to operate the consequent motor control. All working receptors record and communicate discreetly. Each sensory system performs in its highly optimized range of frequencies to integrate data with others and provide adequate execution and timing [81]. Often it has not enough time to obtain and process all kinds of information necessary to execute complex anticipatory strategies. For this, these strategies are often very fast movements, as it happens in numerous disciplines [82]. The effect on the balance of tools compressing the skin (tapes, garments, clothes) is surprising because they may change the perceived feelings, proprioception, and balance, demonstrating, once again, the complexity of life [83–85]. Improvement of hemodynamic, reduction of stress in respiratory, cardiological, and metabolic systems reduced venous system's cross-sectional, activation of tactile mechanoreceptors, attenuation of swinging, and improved proprioceptive feedback couple with an improved equilibrium [86, 87]. Even a relatively low pressure exerted on legs makes athletes able to change and maintain an optimal position, increasing muscles' action and lowering unnecessary movements. It has no measurable negative effects on variables linked to perceived exertion or maximum voluntary isometric strength [87, 88]. Elastic and damping ability are other crucial qualities for athletes influencing their balance and position and even power, maximum strength, and technique. Passive vibrations activate many motor neurons acting on specific muscles [89], implying the risk of premature fatigue [90, 91]. The controversial debate on balance bracelets, with the partially explorable mechanisms supposed to underlie it, constitutes another example of the phenomenon's extraordinary complexity [92–94]. The brain must build absolute movements of head and body in the space for performing elaborated motor tasks. Data coming from proprioception and tactile receptors relate only to the internal environment, that is, relative movements and forces of body segments. Therefore, the balance of the head grouping labyrinths, eyes, and ears and making all these receptors work together must be somewhat privileged [82]. Much research stated the crucial role of head stabilization in different animals: migratory birds use the neck and whole body, athletes try to keep it in as much as possible in a controlled state while jumping or running [82, 95–98]. The labyrinth is an inertial, triaxial sensor equipped with semi-circular canals defining a local reference system without needing external support. It may not properly work under thermal stress, pressure compensation, or diseases [99, 100]. Due to the mix of inertial and gravitational accelerations, all inertial information may be ambiguous. The labyrinths move with the head local reference system to drift, provide illusory sensations, and fail to discriminate external environments' correct orientations. These ambiguities are figured out by combining visual and acoustic information. Vision and sense of sight are decisive [81, 101–105]. Even light sounds made by movements help the sensory feedback. Between actions and sounds, it is established a functional correspondence, so athletes and coaches are used to correlating force–time curves with noise, sometimes understanding complex parameters of movements that cannot be differently analyzed [106–108]. This has important implications for the management, selection, study, and optimization of any sports equipment requiring compromises between handling and precision, stiffness and elasticity, and being evaluated by personal feedback. Athletes are not machines and, in their judgment, may be influenced by age, experience, gender, technique, damping ability, motivation, prejudices, prize, fatigue, placebo effect, and stress. The ability to feel and manage mechanical vibrations relates to skill level and individual characteristics. Dissimilarities are magnified above certain thresholds of frequencies and amplitudes. The ability to use efficient strategies of rhythmic muscle contractions and regulation results in active damping, control of useful DOFs, improved stiffness of the muscle–tendon system, and, finally, a higher efficiency [71]. Sports equipment constrains the athletes and affects their perceptions determining the movement efficiency and capacity to provide accurate technical evaluations. Feelings and perceptions are unique to each body and partially incommunicable. So, the process of linking mechanical characteristics with personal feelings coming from many individuals and skills is a very challenging task [109]. The know-how is widely distributed among different expertise and not always univocal and objectified [110].

4 The Challenges of Biology and Physiology

The living beings are open systems of systems, interlaced networks, complex, looking for but far away from equilibrium, adaptive, time-varying, irreversible, and dissipative. They are multi hierarchical, operating over many scales of space and time, with memory and nonlinear feedback, the meeting point of many weak forces not always all clearly identifiable and none of which usually dominant. There is a complex and unique interaction between internal and external environments and a strong path-dependence unpredictable in advance. Systems of extraordinary complexity such as the brain, microbiome, and immune systems are tied together in the body [111], continuously and dynamically tuned by many bio-oscillators regulated by circadian, hormonal, seasonal, or ovulation cycles [112]. So, they are not graspable by simple static or taxonomic approaches. The biological equilibrium is contingent and temporary, both stable and flexible, enabling it to provide adequate responses to a continuously varying external stimulus [43]. The lack of universal rules to segment living organisms allowing simple causal analysis of their multiple and interlaced functions [113], the impossibility to isolate a single privileged causal level, and the variability in response to any given treatments [114–116] lead to biological relativity [117]. Fractal footprints and attractors are ubiquitously noticeable across many and different scales in neurophysiology, anatomy, physiological, and biomechanical behaviors [118]. Life is doubly unique because of its deterministic and historic-evolutionary nature [119]; the same results are achievable through very different pathways. Due to the great number of involved variables, links, and relations, special analytical techniques are required [120]. Historical-evolutionary mechanisms combined with unpredictable random events lead to unique living systems. So, the generalization of biological laws is difficult. When studying complex systems, there is a strong asymmetry between the inductions possible from available information and what can be explained after something has happened, giving the chance to forecast the future evolution from an actual observable state. Especially in the presence of dynamic and open systems, it is never possible to be sure to have all and really important information and to know the covering laws [121–124]. Looking at regularity and repetitions to predict the future is a very old problem even when faced in a modern perspective because it implicitly supposes that all is written in the past. However, this method needs sufficiently long historical series and the working variables to be known. It assumes that the DOFs are limited, and the system dynamics remain stable over time [125]. There are necessary conditions to win a competition, but the identification of sufficient conditions is a different matter. The way Biology and Physics are closely related is very complex and nonlinear. Retroactions, feedbacks, innumerable interactions, and dynamic processes not statically knowable are at work and difficult to dissect and understand. Simultaneous causes combine with multiple possible solutions. Paths of only a few biochemical reactions can link most pairs of metabolites. Local variations of concentrations could reach the whole network in a very short time. Few molecules take part in many reactions, modeling the network structure, and the experimentally observable functions. An open issue is the application of general properties to single individuals [115]. So, sweeping generalizations are rarely correct, and even when something *usually* occurs this does not mean that it must *always* happen [126]. Athletes play sport. They are living systems, open and dissipative structures looking for homeostasis and allostasis, continually negotiating matter, information and energy between inside and outside, where every chemical, physical, and thermodynamic equilibrium is temporary and contingent [118, 127]. Through hormones and other messengers, physical activity leads to extensive changes in cells, tissues, and organs, challenging the whole-body homeostasis and stimulating integrated, often redundant responses. Multiple and complex cellular networks are involved in this process where the muscle share information with other organs resulting in valuable effects on fitness and performance. Interference effects occur when training simultaneously for endurance and strength; skeletal muscles spread messages between different systems. The Biology of exercise requires an integrative attitude because of the upward and downward multiple and contemporary interactions between cells, genes, molecules, and organs [128, 129]. The integration and mutual adaptation produced during exercise find a powerful conceptualization in psychobiology [130]. As Poincaré already understood guessing the chaos in the apparently simple problem of the three bodies at the end of the nineteenth century, it is impossible to use simple cause–effect reasoning to study a system where numerous interlaced parts are simultaneously at work [131]. Recently, the coordination dynamics perspective has been used to explain the nonlinear interlinkages observed between different human subsystems. Disparate constraints operating at different levels lead to adaptations. This is possible thanks to configurations cooperating at various levels and shows practical applications to sports training [132]. No two bodies are equal, each one having personal and unrepeatable anthropometry and biomechanical characteristics, experience, skills, feelings, highly optimized technical movements adapted, and modeled after years of training, thousands of cycles, and, maybe, even from suffered injuries. Equipment is highly customized and sports-specific, heavily influencing how the athlete perceives the external world and acts. For a long time, biological models have been searched to explain the relationship between training and performance and why some athletes are faster than others [133]. Heterogeneous inputs such as neurotransmitters, hormone and substrate concentrations, ventilation rate, loads and pain, environments, and motivation play a role during exercise while the perception of physical effort is highly personal. Fatigue is an elusive multifactorial concept, unique and partially independent of the specific biological state. To be studied, it requires an interdisciplinary approach considering both mind and body [22, 133–135]. Molecular pathways and gene networks regulate the response to training with the initial conditions. The outcome is individual and should include genetic and epigenetic features, previous training history, and transient functional biological states [136]. The exact quantification of training loads is difficult. Complex open biological systems show interlaced interdependencies and sensitive dependence to initial conditions. Any given single external treatment may generate an unpredictable chain of consequences as it happens in diseases with historical–evolutionary development [22, 136]. Heart rate and blood pressure rise and fall along the days following many inputs, often remaining elevated long after the load is removed. Lactate level has a wide range among different athletes with inter- and intra-individual differences and may depend on the diet or previous muscle glycogen storage; not always a physiological steady state is maintained over time [22, 136, 137]. Different systems are involved in energy production in about the first 100 s after starting a maximal effort. Interestingly, many competitions or part of them fall in this range of time, challenging the possibility of using different strategies, or some combination of these. Activation times, passages between one system and the other, quantity and management of energy substrates, transients and recovery, pacing, and the performance model play a crucial, very complex role in training and race management [138]. The complexity of Biology implies that even new anatomical systems can be theorized. An example is the fascial system representing a continuum of soft, collagen-containing, connective tissue bridging micro- to macro-levels and allowing the integration between other different body systems. Myofascial tissue seems to affect the force transmission between muscles, thus influencing *mechanobiology*, athletic performance, and injury risk [139]. The fascia also works for physiological and metabolic homeostasis encompassing nerves and carrying hormones. It appears to play a key role in body regeneration [140, 141]. Nutrition represents the ideal concluding example for this section. The food comprises more than 26,000 known distinct biochemical components [142], and an ideal sports drink like cow's milk [143] has over 100,000 different molecular species. This is transformed and assimilated with a personal and unique ecosystem:

the microbiome. Food ingestion's order affects the subsequent serum glucose profile and shapes the hormonal profiles [144]. Many variables connected by countless links generate virtually all kinds of outcomes and correlations and deeply challenge our scientific knowledge in the field [145–149].

5 The Sports Environment: Opponents, Equipment, Field of Play, Weather, and Public

How to select the best wheat to optimize a crop yield? The question is not trivial because, in addition to cereal properties under investigation, innumerable external environmental variables such as soil composition, insolation, light, and water reaching the plants can interfere among them, overlapping the genuine differences between cereals. Nothing can be studied neglecting its ecosystem. The standard experimental approach is straightforward: try to make everything as homogeneous as possible and control everything supposed to play a role. It is assumed that only the investigated variables are at work while all the others remain inactive. So, any observed final effect is related only to the different used seeds. When evaluating the effects of a modification or intervention in sports, the same process is theoretically performed. So, all attempts are realized to manipulate only the working variables locking everything else, and it is assumed that all important information is known. Strong assumptions that all other things being equal (*cetertis paribus*) are made [150]. The most important external environmental factors such as heat, wind, altitude, temperature, humidity, visibility, the position of sun and shadow during the competition, vertical planimetry, playing surface, sailing regatta course, type and state of snow and ice have been understood and studied for a long time [151–153]. Athletes are trained to deal with these conditions [154]. Unfortunately, in reality, it is very difficult to solve these issues scientifically. Ronald A Fisher—a father of modern statistics—was aware that uncertainty and error are inevitable and consequently proposed to work with a *rigorously specified uncertainty* for making sense of the world and to acquire reliable empirical knowledge [155]. To his seminal ideas, we owe the concepts of randomization and double-blind in the sciences of life. However, in sport, it is not always possible to operate in this way. The outdoor field changes his status during the race. Environmental impact is crucial even in disciplines held indoor: temperature, humidity, light, and noise can affect the performance. How is it possible to randomize it? Quality and temperature of air, light and shadow, crowd presence, and the order in which the athletes are called to play a key role in the outcome [156, 157]. It is practically impossible to fully keep stable and equal for all athletes a track in a marathon, race walking, sailing regatta, downhill and cross-country ski, or bobsleigh and luge. Even the lanes of an athletics track, rowing basin, or swimming pool are not exactly equivalent to each other, so that they are assigned by merit or lot. The ice of the track deteriorates with the passage of blades, the wind changes direction, and the snow varies with sun exposure, daytime, and season. Bioclimatology heavily impacts the perception of fatigue; the light and the background color influence the aiming in the pentathlon, biathlon, and archery; the poor visibility affects the balance of the skier, as well as the waves, do with rowers, sailors, and surfers. During a test, it is normal that temperature and wind change. Furthermore, the number of competitions in a season is limited. For various reasons, the best athlete or team does not always win [31], and confirmations or rebuttals are not always possible due to different environmental conditions and races' unicity. It is difficult to fully perform the procedures orthodoxly required for tests in biosciences: randomizations, control groups, and double-blind protocols. Sometimes the very foundations of controlled experimentation are critical. In fact, often, the number of available top-level athletes is small, random control groups are impossible for practical or ethical reasons, and blind or placebo tests are unrealistic because everyone knows what he is doing and the equipment being tested. Athletes, coaches, and researchers are humans and have expectations and prejudices. Economic or other factors may be present, and not all boundary variables can be strictly controlled. It cannot be excluded that placebo and nocebo effects are at work [158]. A working strategy is inductive. The classic example is the Evidence-Based Medicine, where systematic research extracts the best available general evidence to be integrated with individual know-hows. A lot of data and proper statistics are required. This approach provides useful information in many areas, including injury prevention and treatment, stretching, and physical examination reliability [159]. Huge databases of competition data are widely available for sports analytics, and appropriate statistical tools extract general and transversal information looking for the best key performance indicators. Regression analysis is utilized when variables are continuous such as the official timings. Multiple linear regression has been tested and tuned: several variables are supposed independent, calibrated with proper coefficients, added together, and then used as input in systems of equations. The purpose is to estimate a predictive variable [48, 156, 160, 161]. Both fixed and random effects participate in forming these models, which can be organized as hierarchic or multi-level architecture. These characteristics tune models as linear, additive, and mixed. The observed values might vary from predicted values, and this difference permits evaluations and adjustments [50, 121, 162]. These tools are applied in many sports and for circumstances like the "home advantage" [163]. The outcomes are potentially interesting for researchers looking for all conceivable reasons of variability in test or race. For example, it is possible to investigate the effects of ground, snow type, and altitude separately from the bio-variability of human performance, so finding more or less remarkable stability of performances of top athletes and the eventual role of chance. Similar analysis and results are found in rowing [164], while interestingly, the predictability of performance is high when considering all athletes but low for top ones [51]. Other challenging results regarding the effect of the environment on performance come from sliding sports where top athletes' performance is variable and difficulty of different tracks, effects of ice softening and degradation, home advantage, and unpredictability of race outcomes are assessed [49, 165]. The multiplicity of applied pieces of knowledge and approaches to sports and physical activity is proven by the great variety of tools used for analysis [20, 64, 160, 166–171]. Considering all the competitions of 20 different Winter Olympics editions and excluding sports with subjective and esthetic judgments, home advantages were found only for few disciplines and not for others, contrarily to popular beliefs [161, 172]. Finally, it is worth mentioning that more or less consciously, other athletes, opponents, contestants, and competitors are an important environmental and unpredictable factor [173–175].

6 Testing: Where Integrated Competencies Meet

Testing and analyzing sports' performance is a matter of paramount importance but, at the same time, often a challenging issue. It is a sort of fractal process: each question opens other questions, and so on. The possibility to follow all normal procedures shows some limitations: careful assessments of internal and external validity should be made, the used surrogate endpoints must be properly validated and this is not always possible, laboratory and tests usually are more accurate and simpler to perform than those carried out in the field but not always representative, the number of possible tests with high-level athletes is often limited [64, 171, 176], the use of instruments on filed and worn on the body it is not always easy and may disturb the action, data are incomplete or cannot be used due to their poor quality or for technical issues, the time necessary for post-processing is long and not compatible with the needs of coaches and athletes [177]. In outdoor sports characterized by environmental variables, it is not always possible to evaluate the impact of field conditions on the outcome. Sometimes there is no way to make the competition field and track uniform in space and time and consequently fair for all athletes. Thus, in alpine skiing, the race numbers are drawn by lot between different merit groups, and the descent order of best athletes is reversed in the second heat. Enormous forces and torques arise in a fraction of seconds and are transmitted to external constraints across very small surfaces, thus generating transients and enormous pressures. External variability has multiple origins, and his measure and control are practically implausible over certain thresholds. In turn, there is an internal variability related to biological mechanisms. The human body continuously sets itself to respond to internal and external environmental changes: heart rate, body temperature equilibrium, flexibility and stiffness, production of force, and power are dynamically tuned. Perceived exertion of fatigue, circadian fluctuations, menstrual cycle, and jet lag may affect the results of the test and the evaluation of performance as well as of any biomarkers [112, 178–180]. So, the characteristics of the human systems entail special care to be studied [63]. Skeletal muscles play the role of agonists and antagonists at the same time, so that a motor task can be obtained in different ways. Available data for muscle activation are few and discrete as it occurs with electromyography. So, observations are inevitably limited to the "surface," the iceberg's tip, and the scientists try to infer the underlying processes [181]. This makes impossible to define the motor patterns univocally, so that some authors theorized the existence of a "motor equivalence," the possibility to use different coordinative movement patterns to produce similar outcomes functionally. Consequently, all different working subsystems are not experimentally separable, and some hidden phenomena are only speculated. To verify causal relationships, an ideal experiment should focus on one variable assuming all other constants and checking the effect on the outcome. A certain number of trials are needed to gain the reliability as a function of the selected parameter's type and consistency. In sports, this is often operatively difficult, if not impossible, for practical reasons [64] or because the movements under investigations are not simply graspable by single kinematics measures [59]. A limited number of feasible tests, gender bias because most studies are performed on males [64, 171], challenges in defining the proper predictors able to explain a performance [182] are all complications in applied sports research. Extraordinary issues arise when studying the interaction between crew members' coordination and a boat's performance in the presence of external and meteorological variables as it happens in rowing and sailing [183, 184]. When analyzing single sections of a race or a portion of data, there is the risk of losing critical information because each part of a performance is also a function of previous ones. Studying only a section might not be enough to understand and explain the global result [185, 186]. The Physics of sport is of extraordinary complexity when closely studied [187]. A crucial property of biological tissues is viscoelasticity, that is, the continuous deformation with time under constant loads. The stress decreases, and there is hysteresis [13]. As seen before, the human body is not rigid, has internal dynamics, redundant DOFs, not uniform density distribution, soft tissues, and wobbling masses. The external and internal mechanical information is read by receptors of the athletes and dynamically processed in a very short time by the Central Nervous System to provide suitable output, making stiffer or softer the whole body or its parts [82, 188]. It happens thanks to complex and partially unexplored feedback and control. For all these reasons, it is difficult to simplify and find solutions to differential equations systems. In sport, it is a well-known fact that not all the ideal solutions provided by Physics are applicable. Even well-trained top athletes are not able to use them on the field profitably [189]. This shows how it is necessary to consider biology's role, and everything is functional to humans. Adequate tools and models are required [9, 74, 160, 168–170], remembering the elusive nature of human features not all directly measurable and conceptualizable. Therefore, hypothetical and multifactorial constructs are used [190]. Even if close to the top performance, some limit configurations cannot be directly explored because of their instability and dangerousness. The definition and use of Key Performance Indicators (KPI) are becoming more and more widespread [191]. But it is necessary to remember that KPI refers to theoretical constructs. All models should be validated, wondering if simplifying and slicing the reality in preselected variables something important is being lost. This again emphasizes the importance of the integrated viewpoint. A theoretical approach allows sensitivity analysis, investigates the relationships between variables, and comes closer to understanding the causes than experimental studies looking for correlations. So, models and simulations are increasingly used to perform calculations, standardize the boundary conditions, and carry parametric tests without risks and at low cost. The optimum research suggests and hierarchies the most profitable development direction and drives towards appropriate experimental data acquisition. This process helps to adjust the simulations to adhere more and more to reality [192]. Unlike what happens in Formula 1 or in sailing America's Cup, this is difficult in other disciplines where the human body is the largest and heaviest part of the system. So, inertial, postural, muscular, kinematical, athletes' dynamical characteristics must be considered with related ability in motor control [193, 194]. The simulations of athletes, equipment, environment, and opponents are limited by the systems' biological and complex nature with many DOFs, and the internal dynamic is not fully defined. Body geometry reconstruction [195], individual features, muscle models, wobbling masse representing soft tissue, rigidity and viscoelasticity, inverse dynamics, interface with external surfaces, and inertial parameters are examples of challenging issues. The system is path-dependent, self-adapting, continuously fatiguing and recovering, characterized by continuous accumulation and processing of memory, and exceptionally sensitive [6, 196]. So, often broad generalizations are unserviceable, and it is more effective to aim at specific improvements of single top athletes. Contingent and highly customized answers are provided for technique, training, injury prevention, and other details limiting the performance. Finally, a key important role is played by the competition formats, which often, even in the Olympic Games, are intentionally designed to shuffle the cards making the most exciting and unpredictable results but sometimes penalizing the most deserving athletes in favor of the show.

7 Conclusion

Western thought history is characterized by the attempt to tie universal to particular, multiple to unitary, necessary to contingent, ideal to real, objective to subjective, and absolute to relative. Modern Western science was born and developed thanks to the obsession for unbiased measurements. Plato stated that the world of ideas is objective and general. In contrast, life's world is subjective and singular because body and senses provide only personal, contingent, and deceptive sensations. In *the Assayer*, Galileo Galilei claimed that heat, color, flavor, and smell are not objective, *essential* qualities of the bodies but subjective, *sensible* impressions due to our organs. He concluded that *only what can be counted is important: we must measure what is measurable, and make measurable what is not*. In this way, Galilei paved the way for modern redutionist science. Western Medicine has been thinking for at least 25 centuries about jumping between classes' properties to those of individuals. Descartes objectified the body, making mechanism the organism, a sum of parts to analyze using Physics' approaches like dynamics or hydraulics, so opening the possibility to use a Newtonian method to study the living beings and leading to the foundation of modern Medicine and Bio-sciences. The body includes organs, each with objective and measurable features studied by different specialists and producing separated branches: cardiology, orthopedics, ophthalmology, pulmonology, immunology, dermatology, surgery, and so on. The discovery of pathogens whit the correspondence of a bacterium with a disease at the end of the XIX century, and modern genetics in more recent times, further strengthened the reductionist approach where each discipline generates its own knowledge, detached from the others. In the hospital, the sick person becomes a carrier of organs or pathologies and is referred to specialists and diagnostics based on specific symptoms. This anatomical-functional taxonomy has brought extraordinary results for all to see but moved away from the holistic knowledge dividing it into competencies and sectors. The same happened in sports sciences shaped on life sciences: physiology, sports medicine, biomechanics, performance and match analysis, analytics, training, strength, and conditioning, nutrition, vision, psychology, talent identification, coaching, sports equipment and technology research, the study of technical and tactical demands, analysis of Key Performance Indicators, team management, performance planning, neurosciences, and Big Data are all generally recognized as disciplines to be studied for their contribution to the final outcome but hardly tied together to paint an overall picture. In modern sport, this approach has led to enormous progress, as shown by human performances' evolution over time. However, all these different pieces of knowledge, far from being incompatible, require a deep integration to work properly. Sometimes science is the hostage of its method, the mathematical-quantitative process adopted to produce ideal objects used to interpret and comprehend the real world. Considering the athlete and the sports performance in isolation, the purpose, and intention that make each gesture an adequate response to a given situation is lost. The performance is not the product of pre-existing structures but rather the result of unpredictable choices made in a continuous adaptation of the external and internal worlds. As Antonio Machado poetically warned: *there is no way. Walking makes the way*.Having highlighted in the introduction the importance of *integration*, now let briefly dwell on the term *knowledge*. In sports, the know-how is widely distributed among different expertise such as coach, trainer, physiotherapist, analyst, medical doctor, engineer, mechanic, ski man, sports equipment manufacturer, and—ultimately—the athlete. This know-how is not always objectified. It is impossible to teach someone to swim or to ride a bike only in words. Sensations, balance, feedback, and execution technique are strictly personal for each of us and can be transmitted and learned only through complex processes of examples, imitations, trials, and errors. In the 1950s, it was introduced the concept of *tacit knowledge*, the kind of implicit, non-codified knowledge we are not aware of and therefore unable to explicit [110]. It is indefinable know-how between Science and Art based on complex interactions of cognition and memory [197], difficult to transfer but—nevertheless—possibly resulting in significant competitive advantages [198]. Everyone involved in sports should challenge these issues minding that they cannot be separated from each other beyond a certain level. Here is the integrated approach's beating heart: a diversified group capable of disparate and dynamical answers. In modern sport, different professionals' competencies are required to support the Athlete and the Coach, and a global view should be recommended. It is not always possible to understand, calculate, simulate, predict, and control all that happens. Limitations lay on different fundamental reasons: insufficient knowledge of governing laws and initial/boundary conditions; poor understanding of variables distribution and the underlying dynamics; the nature of the system [3, 29]; too many DOFs [125, 199]; not separability between observer and observed; all other things being equal assumptions [150]; historical-evolutionary nature combined with random events; an inextricable mixture of chaos and noise. Complex living organisms show all these features and are open, challenging issues [200]. The study of the emergent behavior at different scales may be a target for the future, as well as the comprehension of how the system dynamics are governed by the constraints acting at various orders and levels. Space and time patterns may exhibit corresponding but, at the same time, different behaviors. As it happens in other biological groups featured by dynamic order emerging from physical principles, random perturbations, and simple rules mixed, team sport performance analysis entails complex interactions of deterministic and evolutionary behavior, leading to still an unsolved problem [16, 132]. Sport embraces many different and pluri-contextual pieces of knowledge, but a Grand Unified Theory of sports performance is still not available. Some researchers even doubt the possibility of achieving unified theories in complex adaptive systems [201]. Frontiers remain to be explored [57], demonstrating, once again, the key role of knowledge integration and why the sport Sciences are an extraordinary field of play.

Integrated Sport Sciences in 2050

According to a famous and provocative definition, playing a game is *a voluntary attempt to overcome unnecessary obstacles according rules to forbid the use of the most efficient means* [202]. Adding the need of body movements, for Sport is the same. So, room for improvements will always exist because sport is made of athletes, opponents, rules, equipment, environments, and records. The formers are, at the same time, the less and the more variable element. The less because in the next years the human body will reasonably remain the same, the more because developments in fields such as neurosciences, artificial intelligence, training, food, and applied technology are unpredictable. Sports rules, equipment and records will depend on how the show and the business will develop. The inclusion of eSports is a challenge for the future. But beyond certain limits it is not possible to predict the course of open, adaptive, and complex systems. And Sport is done by women and men with their emotions, passions, weaknesses. So, it will change and evolve to likely remain that exciting and unpredictable phenomenon we all know and love.

Core Messages

- Due to its unique characteristics, sport represents an extraordinary challenge for the limits of human physical performance and scientific knowledge.
- The top performance is a multidimensional construct at the intersection between Physics, Chemistry, and Biology, so a multidisciplinary and integrated approach is necessary.

- The uniqueness of each body and the distribution of sports know-how among diverse expertise make the knowledge partially implicit, non-codified, and non-transmissible.
- In open, complex, adaptive systems, the relationships between parts and outcome can be speculated within certain assumptions, and sometimes it is only possible to set intermediate, surrogate endpoints to be adequately validated.
- Science of complexity and networks, integrated biology, Big Data, models, and modern approaches brought several answers to old questions in sports and opened new challenging frontiers.

References

1. Sands WA, McNeal JR, Stone MH (2005) Plaudits and pitfalls in studying elite athletes. Percept Mot Skills 100(1):22–24
2. Loland S, McNamee M (2017) Philosophical reflections on the mission of the European College of sport science: challenges and opportunities. Eur J Sport Sci 17(1):63–69. https://doi.org/10.1080/17461391.2016.1210238
3. Caprara S, Vulpiani A (2016) Chaos and stochastic models in physics: ontic and epistemic aspects. In: Models and inferences in science. Springer, Berlin, pp 133–146
4. Firth WJ (1991) Chaos—predicting the unpredictable. BMJ 303(6817):1565. https://doi.org/10.1136/bmj.303.6817.1565
5. Taga G (1995) A model of the neuromusculoskeletal system for human locomotion. I. Emergence of basic gait. Biol Cybern 73(2):97–111. https://doi.org/10.1007/BF00204048
6. Taga G (1995) A model of the neuromusculoskeletal system for human locomotion. II. Real-time adaptability under various constraints. Biol Cybern. 73(2):113–21. https://doi.org/10.1007/BF00204049
7. Knudson DV, Morrison CS (2002) Qualitative analysis of human movement. Human Kinetics, pp 73–189
8. Fleishman EA, Quaintance MK, Broedling LA (1984) Taxonomies of human performance: the description of human tasks. Academic Press INC
9. Troje NF (2002) Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. J Vis 2(5):371–387. https://doi.org/10.1167/2.5.2
10. Hughes MD, Bartlett RM (2002) The use of performance indicators in performance analysis. J Sports Sci 20:739–754
11. Hughes M (2004) Notational analysis—a mathematical perspective. Int J Perf Anal Sport 4 (2):97–139
12. Hall D, James D, Marsden N (2012) Marginal gains: Olympic lessons in high performance for organisations. HR Bullet: Res Pract 7(2):9–13
13. Bartlett R (1999) Sports biomechanics: reducing injury and improving performance. Routledge, Abingdon
14. Yeadon MR (2004) What are the limitations of experimental and theoretical approaches in sports biomechanics? In: McNamee (ed) Philosophy and the sciences of exercise, health and sport. Critical perspectives on research methods, Chap 7. Routledge, Abingdon, pp 126–130

15. Hosokawa K, Sato Y, Sakata T (2006) Study on modification of ski referring characteristic of ski turn (change of width and shoe center). In: The engineering of sport, vol 6. Springer, Berlin, pp 213–218
16. Balague N, Torrents C, Hristovski R, Kelso JAS (2016) Sport science integration: An evolutionary synthesis. Eu J Sport Sci
17. Hrysomallis C (2011) Balance ability and athletic performance. Sports Med 41(3):221–232. https://doi.org/10.2165/11538560-000000000-00000
18. Suchomel TJ, Nimphius S, Stone MH (2016) The importance of muscular strength in athletic performance. Sports Med 46(10):1419–1449. https://doi.org/10.1007/s40279-016-0486-0
19. Seiler S (2015) 150 years of rowing faster: what are the sources of more and more speed? BMC Sports Sci Med Rehabil 7(Suppl 1):O12
20. Hopkins WG, Hawley JA, Burke LM (1999) Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc 31(3):472–485. https://doi.org/10.1097/00005768-199903000-00018
21. Noakes TD (2000) Physiological models to understand exercise fatigue and the adaptations that predict or enhance athletic performance. Review Article. Scan J Med Sci Sports 10 (3):123–145. https://doi.org/10.1034/j.1600-0838.2000.010003123.x
22. Borresen J, Lambert MI (2009) The quantification of training load, the training response and the effect on performance. Sports Med 39(9):779–795. https://doi.org/10.2165/11317780-000000000-00000
23. Halperin I, Vigotsky AD, Foster C, Pyne DB (2017) Strengthening the practice of exercise and sport-science research. Int J Sports Phys Perf 13(2):127–134. https://doi.org/10.1123/ijspp.2017-0322
24. Tweedy SM, Vanlandewijck YC (2011) International Paralympic Committee position stand-background and scientific principles of classification in Paralympic sport. Br J Sports Med 45(4):259–269. https://doi.org/10.1136/bjsm.2009.065060
25. Miah A (2006) Rethinking enhancement in Sport. Ann N Y Acad Sci 1093:301–20. https://doi.org/10.1196/annals.1382.020
26. James D (2010) The ethics of using engineering to enhance athletic performance. Proc Eng 2 (2):3405–3410. https://doi.org/10.1016/j.proeng.2010.04.165
27. Kram R, Grabowski AM, McGowan CP, Brown MB et al (2010) Last word on point: counterpoint: artificial limbs do/do not make artificially fast running speeds possible. J Appl Physiol 108(4):1020–1020
28. Vv. Aa (2013) Challenges in irreproducible research. Specials & supplements archive. Nature 456
29. Aicinena S (2013) The impact of chaos, complexity, and luck on coaching success. Int J Soc Sci Educ 3(3). ISSN: 2223-4934
30. Sire C, Redner S (2009) Understanding baseball team standings and streaks. Eur Phys J 67 (3):473–481
31. Kuper S, Szymanski S (2014) Soccernomics: why England loses, why Spain, Germany, and Brazil win, and why the US, Japan, Australia and even Iraq are destined to become the kings of the world's most popular Sport. Nation Books
32. Heuberger JA, Cohen Tervaert JM, Schepers FM, Vliegenthart AD, Cohen AF (2013) Erythropoietin doping in cycling: lack of evidence for efficacy and a negative risk–benefit. Br J Clin Pharmacol 75(6):1406–21. https://doi.org/10.1111/bcp.12034
33. Ellis GF (2008) On the nature of causation in complex systems. Trans Royal Soc South Africa 63(1):69–84
34. Ward-Smith A, Clements D (1982) Experimental determination of aerodynamic characteristics of ski jumpers. Aer J 86:384–391
35. Muller W, Platzer D, Schmolzer B (1996) Dynamics of human flight on skis: improvements in safety and fairness in ski-jumping. J Biomech 29(8):1061–8. https://doi.org/10.1016/0021-9290(95)00169-7

36. Muller E, Schwameder H (2003) Biomechanical aspects of new techniques in alpine skiing and ski-jumping. J Sports Sci 21(9):679–92. https://doi.org/10.1080/0264041031000140284
37. Bernstein NA (1967) The coordination and regulation of movement. Pergamon Press
38. Latash ML (2008) Synergy. Oxford University Press
39. en.wikipedia.org/wiki/Equifinality
40. Kiely J (2018) (2018) Periodization theory: confronting an inconvenient truth. Sports Med 48(4):753–764. https://doi.org/10.1007/s40279-017-0823-y
41. Ajemian R, D'Ausilio A, Moorman H, Bizzi E (2010) Why professional athletes need a prolonged period of warm-up and other peculiarities of human motor learning. J Mot Behav 42(6):381–8. https://doi.org/10.1080/00222895.2010.528262
42. Hatze H (1986) Motion variability. Its definition, quantification and origin. J Mot Behav 18 (1):5–16. https://doi.org/10.1080/00222895.1986.10735368
43. Davids K, Glazier P, Araujo D, Bartlett R (2003) Movement systems as dynamical systems. Sports Med 33(4):245–260. https://doi.org/10.2165/00007256-200333040-00001
44. Davids K, Hristovski R, Araújo D, Serre NB, Button C, Passos P (eds) (2013) Complex systems in sport. Routledge, Abingdon
45. Bartlett R, Wheat J, Robins M (2007) Is movement variability important for sports biomechanists? Sports Biomech 6(2):224–243. https://doi.org/10.1080/14763140701322994
46. Preatoni E, Hamill J, Harrison AJ, Hayes K, Van Emmerik RE, Wilson C, Rodano R (2013) Movement variability and skills monitoring in sports. Sports Biom 12(2):69–92. https://doi.org/10.1080/14763141.2012.738700
47. Hiley MJ, Zuevsky VV, Yeadon MR (2013) Is skilled technique characterized by high or low variability? An analysis of high bar giant circles. Hum Mov Sci 32(1):171–180. https://doi.org/10.1016/j.humov.2012.11.007
48. Bruggemann GP, Morlock M, Zatsiorsky VM (1997) Analysis of the Bobsled and Men's luge events at the XVII Olympic Winter Games in Lillehammer. J App Biomed 13(1):98–108
49. Bullock N, Hopkins WG, Martin DT, Marino FE (2009) Characteristics of performance in skeleton World Cup races. J Sports Sci 27(4):367–372. https://doi.org/10.1080/02640410802613425
50. Hopkins WG, Hume PA, Hollings SC, Hamlin MJ, Spencer M, Malcata RM, Smith TB, Quarrie KL (2013) Mixed linear modelling: a very short introduction. Sportscience 17 July sportsci.org/2013/inbrief.htm#mixed
51. Malcata RM, Hopkins WG (2014) Variability of competitive performance of elite athletes: a systematic review. Sports Med 44(12):1763–74. https://doi.org/10.1007/s40279-014-0239-x
52. Loland S (1992) The mechanics and meaning of alpine skiing: Methodological and epistemological notes on the study of sport technique. J Phil Sport 19(1):55–77
53. Lees A (2002) Technique analysis in sports: a critical review. J Sports Sci 20(10):813–828. https://doi.org/10.1080/026404102320675657
54. Warmenhoven J, Cobley S, Draper C, Smith R (2018) Over 50 years of researching force profiles in rowing: what do we know? Sports Med 48(12):2703–2714. https://doi.org/10.1007/s40279-018-0992-3
55. Yeadon MR, King MA (2017) Computer simulation modelling in Sport. In: Biomechanical evaluation of movement in Sport and exercise. Routledge, Abingdon, pp 176–205
56. Newell KM (1986) Constraints on the development of coordination. In: Wade MG, Whiting HTA (eds) Motor development in children: aspects of coordination and control, vol 34. Martinus Nijhoff, Dordrecht, pp 341–360
57. Vv. Aa (2017) Towards a grand unified theory of sports performance—special Issue. Hum Mov Sci 56(Pt A):139–156. https://doi.org/10.1016/j.humov.2015.08.001
58. Charles JD, Bejan A (2009) The evolution of speed, size and shape in modern athletics. J Exp Biol 212(15):2419–2425. https://doi.org/10.1242/jeb.031161

59. Guignard B, Rouard A, Chollet D, Bonifazi M, Dalla Vedova D, Seifert L (2017) Perception and action in swimming: effects of aquatic environment on upper limb inter-segmental coordination. Hum Mov Sci 55:240–254. https://doi.org/10.1016/j.humov.2017.08.003
60. West GB (2017) Scale: the universal laws of growth, innovation, sustainability, and the pace of life in organisms, cities, economies, and companies. Penguin Press, New York
61. Federolf P, Reid R, Gilgien M, Haugen P, Smith G (2014) The application of principal component analysis to quantify technique in sports. Scan J Med Sci Sports 24(3):491–499. https://doi.org/10.1111/j.1600-0838.2012.01455.x
62. Warmenhoven J, Cobley S, Draper C et al (2019) Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing. Sports Biomech 18(3):317–341. https://doi.org/10.1080/14763141.2017.1392594
63. Hopkins WG, Marshall SW, Batterham AM, Hanin J (2009) Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Ex 41(1):3–12. https://doi.org/10.1249/MSS.0b013e31818cb278
64. Mullineaux DR, Bartlett RM, Bennett S (2001) Research design and statistics in biomechanics and motor control. J Sports Sci 19(10):739–760. https://doi.org/10.1080/026404101317015410
65. Salo A, Grimshaw PN, Viitasalo JT (1997) Reliability of variables in the kinematic analysis of sprint hurdles. Med Sci Sports Exercise 29(3):383–389. https://doi.org/10.1097/00005768-199703000-00014
66. Glazier PS, Davids K (2009) Constraints on the complete optimization of human motion. Sports Med 39(1):15–28. https://doi.org/10.2165/00007256-200939010-00002
67. Glazier PS, Mehdizadeh S (2019) In search of sports biomechanics' Holy Grail: can athlete-specific optimum sports techniques be identified? J Biomech 94:1–4. https://doi.org/10.1016/j.jbiomech.2019.07.044
68. Natri A, Beynnon BD, Ettlinger CF, Johnson RJ, Shealy JE (1999) Alpine ski bindings and injuries. Sports Med. PMID: 10461711
69. Burtscher M, Gatterer H, Flatz M, Sommersacher R, Woldrich T, Nachbauer W (2008) Effects of modern ski equipment on the overall injury rate and the pattern of injury location in Alpine skiing. Clin J Sport Med. 18(4):355–7. https://doi.org/10.1097/MJT.0b013e31815fd0fe
70. Mester J (1996) Movement regulation in alpine skiing. In: Müller E, Schwameder H, Kornexl E, Raschner C (eds) Science and skiing. E&FN Spon, pp 333–348
71. Haas CT, Simon C, Schmidtbleicher D (1998) Simulation of vibrations in alpine skiing. In: ISBS-conference proceedings archive, vol 1, No 1
72. Nemec B, Kugovnik O, Supej M (2001) Influence of the ski side cut on vibrations in alpine skiing. Sci Skiing II:232–241
73. Davlin CD (2004) Dynamic balance in high level athletes. Percept Mot Skills 98(3 Pt 2):1171–6. https://doi.org/10.2466/pms.98.3c.1171-1176
74. Danion F, Latash ML (2011) Motor control. Theories, experiments and applications. Oxford University Press, Oxford
75. Schwarz US (2017) Mechanobiology by the numbers: a close relationship between biology and physics. Nat Rev Mol Cell Biol 18(12):711–712. https://doi.org/10.1038/nrm.2017.109
76. Haas CT, Turbanski S, Schmidtbleicher D (2005) Improved sensorimotor control is not connected with improved proprioception. Br J Sports Med 39(Suppl 1):6, 388
77. Moberg E (1983) The role of cutaneous afferents in position sense, kinaesthesia, and motor function of the hand. Brain 106 (Pt 1):1–19. https://doi.org/10.1093/brain/106.1.1
78. Collins DF, Prochazka A (1996) Movement illusions evoked by ensemble cutaneous input from the dorsum of the human hand. J Physiol 496:857–71. https://doi.org/10.1113/jphysiol.1996.sp021733
79. Spelmezan D, Jacobs M, Hilgers A, Borchers J (2009) Tactile motion instructions for physical activities. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 2243–2252

80. Llinás R (1991) The noncontinuous nature of movement execution. In: Humphrey DR, Freund HJ (eds) Motor control: concepts and Issues. Wiley, Hoboken, pp 223–242
81. Redfern MS, Yardley L, Bronstein AM (2001) Visual influences on balance. J Anxiety Disorders 15(1):81–94
82. Berthoz A (1997) Le sens du movement. Ed Odile
83. Simoneau GG, Degner RM, Kramper CA, Kittleson KH (1997) Changes in ankle joint proprioception resulting from strips of athletic tape applied over the skin. J Athl Train 32 (2):141–147 PMID: 16558444
84. Ibegbuna V, Delis KT, Nicolaides AN, Aina O (2003) Effect of elastic compression stockings on venous hemodynamics during walking. J Vasc Surg 37(2):420–425 PMID: 12563216 Clinical Trial
85. Perrey S (2008) Compression garments: evidence for their physiological effects. In: Estivalet (ed) The engineering of sport 7, vol 2 P208. Springer, Berlin
86. Schaff P, Hauser W (1990) Ski boots versus the knee joint. What produces the forward leaning position of the ski boot? Sportverletz Sportschaden 4(1):1–13
87. Sperlich B, Born DP, Swarén M, Kilian Y, Geesmann B, Kohl-Bareis M, Holmberg HC (2013) Is leg compression beneficial for alpine skiers? BMC Sports Sci Med Rehab 5(1):18. https://doi.org/10.1186/2052-1847-5-18
88. Born DP, Sperlich B, Holmberg HC (2013) Bringing light into the dark: effects of compression clothing on performance and recovery. Int J Sports Physiol Perform 8(1):4–18. https://doi.org/10.1123/ijspp.8.1.4
89. Cardinale M, Lim J (2003) Electromyography activity of vastus lateralis muscle during whole-body vibrations of different frequencies. J Strength Cond Res. PMID: 12930196 Clinical Trial
90. Burke D, Hagbarth KE, Lofstedt L, Wallin BG (1976) The responses of human muscle spindle endings to vibration of non-contracting muscles. J Physiol 261(3):673–93. https://doi.org/10.1113/jphysiol.1976.sp011580
91. Perrey S (2008) Compression garments: evidence for their physiological effects. In: Estivalet M, Brisson P (eds) The engineering of sport 7, vol 2. Springer, Berlin, pp 319–328
92. Bratton RL, Montero DP, Adams KS, Novas MA, Mckay TC, Hall LJ, Foust JG, Mueller MB, O'Brien PC, Atkinson EJ, Maurer MS (2002) Effect of "Ionized" Wrist Bracelets on musculoskeletal pain: a randomized, double-blind, placebo-controlled trial. Mayo Clin Proc 77:1164–1168
93. Bringman E, Kimura C, Philip S (2011) Effects of an ionic bracelet on physical, cognitive, and integrative tasks. Pacific Northwest J Undergraduate Res Creative Activities: vol 2, Article 4
94. Porcari J (2011) Can the power balance® bracelet improve balance, flexibility, strength and power? J Sports Sci Med 10(1):230–1. PMCID: PMC3737916
95. Pozzo T, Berthoz A, Lefort L (1990) Head stabilization during various locomotor tasks in humans. Exp Brain Res 82:97–106. PMID: 2257917
96. Pozzo T, Levik Y, Berthoz A (1995) Head and trunk movements in the frontal plane during complex dynamic equilibrium tasks in humans. Exp Brain Res 106(2):327–338. https://doi.org/10.1007/BF00241128
97. Cromwell R, Schurter J, Shelton S, Vora S (2004) Head stabilization strategies in the sagittal plane during locomotor tasks. Physiother Res Int 9(1):33–42. https://doi.org/10.1002/pri.298
98. Armstrong B, McNair P, Taylor D (2008) Head and neck position sense. Sports Med 38 (2):101–117. 0112-1642/08/0002-0101/$48.00/0
99. Nashner LM, Shupert CL, Horak FB, Black FO (1989) Organization of posture controls: an analysis of sensory and mechanical constraints. Prog Brain Res 80:411–8; discussion 395–7. https://doi.org/10.1016/s0079-6123(08)62237-2
100. Bottini G (1995) Modulation of conscious experience by peripheral sensory stimuli. Nature 376(6543):778–81. https://doi.org/10.1038/376778a0

101. Lestienne F, Soechting J, Berthoz A (1977) Postural readjustments induced by linear motion of visual scenes. Exp Brain Res 28(3–4):363–384 PMID: 885185
102. Kerr B, Condon SM, McDonald LA (1985) Cognitive spatial processing and the regulation of posture. J Exp Psychol Hum Percept Perform 11(5):617–22. https://doi.org/10.1037//0096-1523.11.5.617
103. Andersen GJ, Dyre BP (1989) Spatial orientation from optic flow in the central visual field. Percept Psychophys 45(5):453–8. https://doi.org/10.3758/bf03210719
104. Poggio T, Bizzi E (2004) Generalization in vision and motor control. Rev Nature 431 (7010):768–774 PMID: 15483597
105. Fonn D (2011) A special issue on sports vision. Eye Contact Lens 37(3):113. https://doi.org/10.1097/ICL.0b013e31821cf4e
106. Sigrist R, Rauter G, Marchal-Crespo L, Riener R, Wolf P (2015) Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. Exp Brain Res 233(3):909–25. https://doi.org/10.1007/s00221-014-4167-7
107. Bevilacqua F, Boyer EO, Françoise J, Houix O, Susini P, Roby-Brami A, Hanneton S (2016) Sensori-motor learning with movement sonification: perspectives from recent interdisciplinary studies. Front Neurosci 10:385. https://doi.org/10.3389/fnins.2016.00385
108. Schaffert N, Janzen TB, Mattes K, Thaut MH (2019) A review on the relationship between sound and movement in sports and rehabilitation. Front Psychol 10:244. https://doi.org/10.3389/fpsyg.2019.00244
109. Federolf P, Auer M, Fauve M, Lüthi A, Rhyner H (2006) Subjective evaluation of the performance of alpine skis and correlations with mechanical ski properties. Eng Sport 6:287–292
110. Polanyi M (1966) The Tacit dimension. University of Chicago Press, Chicago
111. Fargier P, Collet C, Moran A, Massarelli R (2017) Inter-disciplinarity in sport sciences: the neuroscience example. Eur J Sport Sci 17(1):42–50. https://doi.org/10.1080/17461391.2016.1207710
112. Drinkwater BL (2008) Women in sport. The encyclopaedia of sports medicine: An IOC Medical Commission Publication, vol 8. Wiley, Hoboken
113. Lewontin RC (2001) The triple helix: gene, organism, and environment. Parts and wholes, causes and effects. Harvard University Press, Cap III
114. Ross R, Goodpaster BH, Koch LG et al (2018) Precision exercise medicine: understanding exercise response variability. Br J Sports Med 53(18):1141–1153. https://doi.org/10.1136/bjsports-2018-100328
115. Fisher AJ, Medaglia JD, Jeronimus BF (2018) Lack of group-to-individual generalizability is a threat to human subjects' research. Proc Natl Acad Sci USA 115(27):E6106–E6115
116. Bouchard C, Rankinen T (2001) Individual differences in response to regular physical activity. Med Sci Sports Exerc 33(6 Suppl):S446–51; discussion S452–3. https://doi.org/10.1097/00005768-200106001-00013
117. Noble D (2016) Dance to the tune of life: biological relativity. Cambridge University Press, Cambridge
118. West BJ (2012) Fractal physiology and chaos in medicine, vol 16, 2nd ed. World Scientific, Singapore
119. Monod J (1979) On chance and necessity. Palgrave, London
120. Cheng F, Kovács IA, Barabási AL (2019) Network-based prediction of drug combinations. Nat Commun 10(1):1–11. https://doi.org/10.1038/s41467-019-09186-x
121. Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Ass 88(421):9–25. http://www.jstor.org/stable/2290687
122. Sterman JD, Rykiel EJ, Oreskes N, Belitz K, Shrader-Frechette K (1994) The meaning of models. Science, new series, vol 264, No 5157 April, pp 329–331
123. Oreskes N, Belitz K (2001) Philosophical issues in model assessment. In: Model validation: perspectives in hydrological Science, Chap 3. Wiley, Hoboken, pp 23–41

124. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8):e124. https://doi.org/10.1371/journal.pmed.0020124
125. Cecconi F, Cencini M, Falcioni M, Vulpiani A (2012) Predicting the future from the past: an old problem from a modern perspective. Am J Phys 80(11):1001–1008. https://doi.org/10.1119/1.4746070
126. Mayr E (1997) This is biology: the science of the living world, Chap 4. Harvard University Press, Cambridge
127. von Bertalanffy L (1968) General system theory: foundations, development, applications. George Brazilier
128. Hawley JA, Hargreaves M, Joyner MJ, Zierath (2014) Integrative biology of exercise. Cell Nov 159(4):738–49. https://doi.org/10.1016/j.cell.2014.10.029
129. Goldstein DS (2019) How does homeostasis happen? Integrative physiological, systems biological, and evolutionary perspectives. Am J Physiol Regul Integr Comp Physiol 316(4): R301–R317. https://doi.org/10.1152/ajpregu.00396.2018
130. Marcora S (2019) Psychobiology of fatigue during endurance exercise. In: Meijen C (ed) Endurance performance in sport: psychological theory and interventions. Routledge, Abingdon, pp 15–34
131. Kelso JS (1995) Dynamic patterns—the self-organisation of brain and behaviour. E & MIT Press, Cambridge
132. Balague N, Torrents C, Hristovski R, Davids K, Araújo D (2013) Overview of complex systems in Sport. J Syst Sci Complex 26:4–13. https://doi.org/10.1007/s11424-013-2285-0
133. Noakes TD (2012) Fatigue is a brain-derived emotion that regulates the exercise behavior to ensure the protection of whole-body homeostasis. Front Physiol 3:82. https://doi.org/10.3389/fphys.2012.00082
134. Marcora S (2009) Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. J Appl Phy 106(6):2060–2062. https://doi.org/10.1152/japplphysiol.90378.2008
135. Marcora SM, Staiano W (2010) The limit to exercise tolerance in humans: mind over muscle? Eur J Appl Physiol 109(4):763–770. https://doi.org/10.1007/s00421-010-1418-6
136. Kiely J (2012) Periodization paradigms in the 21st century: evidence-led or tradition-driven? Int J Sports Physiol Perform 7(3):242–50. https://doi.org/10.1123/ijspp.7.3.242
137. Halperin I, Emanuel A (2019) Rating of perceived effort: methodological concerns and future directions. Sports Med 1–9. https://doi.org/10.1007/s40279-019-01229-z
138. Losnegard T, Myklebust H, Hallén J (2012) Anaerobic capacity as a determinant of performance in sprint skiing. Med Sci Sports Exerc 44(4):673–681. https://doi.org/10.1249/MSS.0b013e3182388684
139. Zügel M, Maganaris CN, Wilke J et al (2018) Fascial tissue research in sports medicine: from molecules to tissue adaptation, injury and diagnostics: consensus statement. Br J Sports Med 52(23):1497–1497. https://doi.org/10.1136/bjsports-2018-099308
140. Soh KS, Kang KA, Harrison DK (2012) The primo vascular system: its role in cancer and regeneration. Springer, Berlin
141. Gonzalez CA, Driscoll M, Schleip R et al (2018) Frontiers in fascia research. J Bodyw Mov Ther 22(4):873–880. https://doi.org/10.1016/j.jbmt.2018.09.077
142. Barabási AL, Menichetti G, Loscalzo J (2019) The unmapped chemical complexity of our diet. Nature Food 1–5. https://doi.org/10.1038/s43016-019-0005-1
143. James LJ, Stevenson EJ, Rumbold PL (2019) Hulston CJ (2019) Cow's milk as a post-exercise recovery drink: implications for performance and health. Eur J Sport Sci 19 (1):40–48. https://doi.org/10.1080/17461391.2018.1534989
144. Shukla AP, Iliescu RG, Thomas CE, Aronne LJ (2015) Food order has a significant impact on postprandial glucose and insulin levels. Diabetes Care 38(7):e98–e99.https://doi.org/10.2337/dc15-0429
145. Macdonald I (1983) Nonsense and non-science in nutrition. Proc Nutr Soc 42(3):513–523. https://doi.org/10.1079/pns19830058

146. Schoenfeld JD, Ioannidis JP (2012) Is everything we eat associated with cancer? A systematic cookbook reviews. Am J Clin Nutrit 97(1):127–134. https://doi.org/10.3945/ajcn.112.047142
147. Noakes T (2012) Waterlogged: the serious problem of overhydration in endurance sports. Human Kinetics
148. Ioannidis JP (2013) Implausible results in human nutrition research. BMJ 347:f6698. https://doi.org/10.1136/bmj.f6698
149. Ioannidis JP (2018) The challenge of reforming nutritional epidemiologic research. JAMA 320(10):969–970. https://doi.org/10.1001/jama.2018.11025
150. Cartwright N (1999) The dappled world: a study of the boundaries of science. Cambridge University Press, Cambridge
151. Hill AV (1928) The air-resistance to a runner. Proc Royal Society London 102:380–385
152. Lloyd BB (1967) Theoretical effects of altitude on the equation of motion of a runner. In: Margaria R (ed) Exercise and altitude. Excerpta Medica Foundation, Amsterdam, pp 65–72
153. Hollings SC, Hopkins WG, Hume PA (2012) Environmental and venue-related factors affecting the performance of elite male track athletes. E J Sport Sci 12:3:201–206. https://doi.org/10.1080/17461391.2011.552640
154. Pezzoli A, Baldacci A, Faina M, Dalla Vedova, D Cristofori E (2012) Analysis of navigation pattern in the sport of rowing. In: Proceedings of IX World Congress of performance analysis of sport
155. Marks HM (2003) Rigorous uncertainty: why RA Fisher is important. Int J Epidemiol 32:932–937. https://doi.org/10.1093/ije/dyg288
156. Morlock MM, Zatsiorsky VM (1989) Factors influencing performance in bobsledding: I: influences of the bobsled crew and the environment. Int J Sport Biom 5:208–211. https://doi.org/10.1123/ijsb.5.2.208
157. Kietzig AM, Hatzikiriakos SG, Englezos P (2009) Ice friction: the effects of surface roughness, structure and hydrophobicity. J Appl Phys 106(2):024303
158. Hurst P, Schipof-Godart L, Szabo A (2020) The Placebo and Nocebo effect on sports performance: a systematic review. Euro J Sp Sci 1–14.https://doi.org/10.1080/17461391.2019.1655098
159. MacAuley D, Best MT (2002) Evidence-based sports medicine. BMJ Books
160. Atkinson G, Nevill AM (2001) Selected issues in the design and analysis of sport performance research. J Sports Sci 19(10):811–27. https://doi.org/10.1080/026404101317015447
161. Balmer NJ, Nevill AM, Williams M (2010) Home advantage in the Winter Olympics (1908–1998). J Sports Sci 19(2):129–139. https://doi.org/10.1080/026404101300036334
162. Krzywinski M, Altman N (2015) Multiple linear regression. Nat Methods 12(12):1103–4. https://doi.org/10.1038/nmeth.3665
163. Nevill AM, Holder RL (1999) Home advantage in sport: an overview of studies on the advantage of playing at home. Sports Med 28(4):221–36. https://doi.org/10.2165/00007256-199928040-00001
164. Smith TB, Hopkins WG (2011) Variability and predictability of finals times of elite rowers. Med Sci Sports Exerc 43(11):2155–2160. PubMed doi: https://doi.org/10.1249/MSS.0b013e31821d3f8e
165. Bullock N, Hopkins WG (2009) Methods for tracking athletes' competitive performance in skeleton. J Sports Sci 27(9):937–940. https://doi.org/10.1080/02640410902951386
166. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30 (1):1–15 PMID: 10907753
167. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26(4):217–38. https://doi.org/10.2165/00007256-199826040-00002
168. Gordon D, Robertson E, Caldwell G, Hamill J, Kamen G, Whittlesey S (2004) Research methods in biomechanics. Human Kinetics

169. Gratton C, Jones I (2010) Research methods for sport studies, 2nd edn. Routledge, Abingdon
170. Thomas JR, Nelson JK, Silverman SJ (2011) Research methods in physical activity, 6th edn. Human Kinetics
171. Hébert-Losier K, Supej M, Holmberg HC (2014) Biomechanical factors influencing the performance of elite alpine ski racers. Sports Med 44(4):519–33. https://doi.org/10.1007/s40279-013-0132-z
172. Bray S, Carron A (2017) The home advantage in alpine skiing. Sport Sci Health 13(2):355–363. https://doi.org/10.1007/s11332-017-0365-6
173. Hyland DA (2001) Opponents, contestants, and competitors: the dialectic of Sport. In: Ethics in sport, pp 80–90
174. Greenlees I, Bradley A, Holder T, Thelwell R (2005) The impact of opponents' non-verbal behaviour on the first impressions and outcome expectations of table-tennis players. Psyc Sport Exercise 6(1):103–115
175. Loffing F, Hagemann N, Strauss B (2010) Automated processes in tennis: do left-handed players benefit from the tactical preferences of their opponents? J Sports Sci 28(4):435–43 https://doi.org/10.1080/02640410903536459
176. Virmavirta M, Komi P (1994) Takeoff analysis of a champion ski jumper. J Coaching Sport Sci 1:23–27
177. Sands WA (2008) Measurement issues with elite athletes. Sports Technol 1(2–3):101–104. https://doi.org/10.1002/jst.17
178. Atkinson G, Reilly T (1996) Circadian variation in sports performance. Sports Med 21 (4):292–312. https://doi.org/10.2165/00007256-199621040-00005
179. Drust B, Waterhouse J, Atkinson G, Edwards B, Reilly T (2005) Review: circadian rhythms. In: Sports performance—an update. Chron Int 22(1): 21–44. https://doi.org/10.1081/CBI-200041039
180. Jones TW, Shillabeer BC, Cardinale M (2020) Skin temperature, training load, and subjective muscle soreness in junior endurance athletes: a case study. Int J Sports Physiology Perfo 1(aop), pp 1–4
181. Kelso JS, Holt KG, Flatt AE (1980) The role of proprioception in the perception and control of human movement: toward a theoretical reassessment. Percept Psychophys 28(1):45–52. https://doi.org/10.3758/bf03204314
182. Spörri J, Kröll J, Schwameder H, Müller E (2012) Turn characteristics of a top world class athlete in giant slalom: a case study assessing current performance prediction concepts. Int J Sports Sci Coach 7(4):647–660
183. Cuijpers LS, Passos PJM, Murgia A, Hoogerheide A, Lemmink KM, de Poel HJ (2017) Rocking the boat: does perfect rowing crew synchronization reduce detrimental boat movements? Scand J Med Sci Sports 27(12):1697–1704. https://doi.org/10.1111/sms.12800
184. Terrien E, Huet B, Iachkine P, Saury J (2020) Coordination between crew members on flying multihulls: a case study on a Nacra 17. J Sports Sci Med 19(2):298–308
185. Supej M, Cernigoj M (2006) Relations between different technical and tactical approaches and overall time at men's world cup giant slalom races. Kin Slov 12(1):63–69
186. Brodie MAD, Walmsley A, Page W (2008) Fusion motion capture: a prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing. Sports Technol 1(1):17–28. https://doi.org/10.1002/jst.6
187. Michael L (2015) The physics of sports. McGraw-Hill
188. Jacobs M (2008) Design and recognition of tactile feedback patterns for snowboarding. PhD Thesis, RWTH Aachen University
189. Barrow JD (2010) Rowing and the same-sum problem have their moments. Am J Phys 78:728–732. https://doi.org/10.1119/1.3318808
190. Atkinson G (2002) Sport performance: variable or construct? J Sports Sci 20(4):291–292. https://doi.org/10.1080/026404102753576053
191. Hughes M, Franks I (2008) The essentials of performance analysis: an introduction. Routledge, Abingdon

192. Braghin F, Donzelli M, Melzi S, Sabbioni E (2011) A driver model of a two-man bobsleigh. Sports Eng 13(4):181–193. https://doi.org/10.1007/s12283-011-0066-3
193. Zatsiorsky VM (2000) Biomechanics in sport. Performance enhancement and injury prevention. The encyclopaedia of sports medicine: an IOC medical commission publication, vol 9. Blackwell Science Ltd., Hoboken
194. Latash ML, Lestienne F (2006) Motor control and learning. Springer
195. Cappozzo A, Catani F, Della Croce U, Leardini A (1995) Position and orientation in space of bones during movement: anatomical frame definition and determination. Clin Biomech 10 (4):171–178. https://doi.org/10.1016/0268-0033(95)91394-t
196. Cignetti F, Schena F, Mottet D, Rouard A (2010) A limit-cycle model of leg movements in cross-country skiing and its adjustments with fatigue. Hum Mov Sci 29(4):590–604. https://doi.org/10.1016/j.humov.2010.03.001
197. Nash C, Collins D (2006) Tacit knowledge in expert coaching: science or art? Quest 58 (4):465–477. https://doi.org/10.1080/00336297.2006.10491894
198. Berman SL, Down J, Hill CW (2002) Tacit knowledge as a source of competitive advantage in the National Basketball Association. Ac Manag J 45(1):13–31. https://doi.org/10.5465/3069282
199. West BJ (2006) Where medicine went wrong: rediscovering the path to complexity, vol 11. World Scientific, Singapore
200. Gleick J (2011) Chaos: making a new science. Open Road Media, New York
201. Horgan J (1995) From complexity to perplexity. Can science achieve a unified theory of complex systems? Even at the Santa Fe Institute, some researchers have their doubts. Sci Am 272:104–109. https://www.jstor.org/stable/https://doi.org/10.2307/24980838
202. Suits B (1978) The grasshopper: games, life and utopia. University of Toronto Press

Dario Dalla Vedova , an engineer, works at the sports science Institute of CONI—Italian National Olympic Committee in Rome since 1992. He is involved in the biomechanical evaluation of the Italian National Teams' top athletes and the study on characterization and optimization of sports equipment. He was in the CONI Missions, providing technical support to the Italian team at the Winter Olympic Games of Turin 2006, Sochi 2014, PyeongChang 2018, and the Summer Olympic Games of Rio 2016 and Tokyo 2020. He worked in the CONI—Ferrari project to optimize vehicles, materials, and analytics in Olympic Sports. Dario teaches biomechanics, performance analysis, research methodology, and sports management at the CONI Central Sports School and in courses organized by the Italian National Sports Federations and Universities. He has been the author of national and international publications and book chapters in sports science and biomechanics. He is an invited speaker at national and international conferences.

Single-Valued Neutrosophic Set: An Overview

26

Surapati Pramanik

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

Albert Einstein

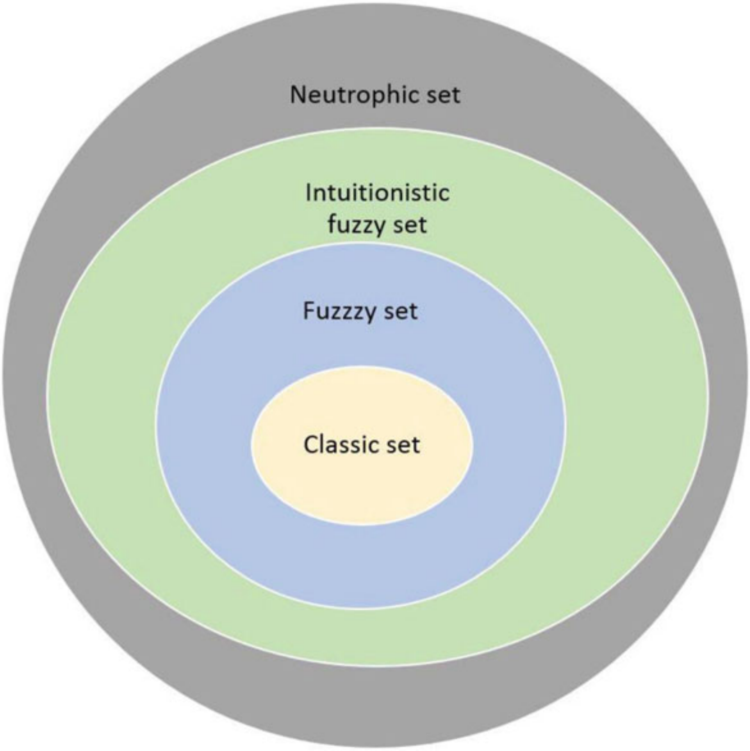
Summary

The purpose of this chapter is to present an overview of neutrosophic sets. Professor Florentin Smarandache defined the neutrosophic set and helped popularize the concept to deal with uncertainty, inconsistency, and indeterminacy common in human existence. The chapter presents the basic definitions of neutrosophic sets, single-valued neutrosophic sets, single-valued neutrosophic numbers, score functions, accuracy and certainty functions, the ranking of neutrosophic numbers, and some extensions of neutrosophic sets. It describes different types of neutrosophic sets and a few examples of their applications in social sciences. The chapter also presents a critical discussion and the future scope of research.

S. Pramanik (✉)
Department of Mathematics, Nandalal Ghosh B.T. College, Panpur, Narayanpur, West Bengal 743126, India

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Kolkata, India

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_26



The relation between classic set, fuzzy set, intuitionistic fuzzy set, and neutrosophic set

The code of this chapter is 01001110 01100101 01110101 01110100 01110010 01101111 01110011 01101111 01110000 01101000 01111001.

1 Introduction

Since Zadeh [1] grounded the Fuzzy Set (FS) in 1965, a plethora of new theories dealing with imprecision and uncertainty has been reported in the fuzzy literature. Neutrosophic Sets (NSs) grounded by Smarandache [2–6] constitute a generalization of FS and Intuitionistic FS (IFS) [7]. While the FS offers the degree of Truth Membership (TM) of an element in a prescribed set, IFS offers both a degree of TM and a degree of False Membership (FM), whereas NS offers a degree of TM, a degree of Indeterminacy Membership (IM), and a degree of FM. In IFS, TM function and FM function are independent, but in FS, FM function is dependent on TM function. In NS, TM function, IM function, and FM function are independent. Smarandache [8] presented the differences between NSs [2] and various extensions of FSs. Single-Valued NS (SVNS) [9–11] is an instance of NS that has a root in the work of Smarandache [2]. Haibin Wang, the first author, has presented the SVNS [9] in the international seminar in Salt Lake City, USA. NSs [2–6] and SVNSs [9–11] have been presented in different seminars and published in different proceedings and journals to draw much attentions from the researchers.

The popularity of NSs [2] gains momentum after the publication of SVNS [11] and the international journal "Neutrosophic Sets and Systems". NSs and SVNSs have been widely used in Multi-Attribute Decision-Making (MADM) [12–21] and Multi-Attribute Group Decision-Making (MAGDM) [22–24]. Interval Neutrosophic Set (INS) [25, 26] has been proposed as an instance of NS. INS is a subclass of NS, and it considers only subunitary intervals of [0, 1].

NSs have drawn much attention from the researchers. Various applications such as NS-based models have been introduced for options market [27], financial market [28], image denoising [29–32], cluster analysis [33], information retrieval [34, 35], love dynamics [36], video tracking [37], fault diagnosis [38], air surveillance [39, 40], and so on.

After 2010, various extensions of NS have been rapidly proposed in the literature. Ye [41] defined the Simplified NS (SNS) in terms of three numbers in [0, 1]. SNS is a subclass of NS. SNS includes of an INS and an SVNS. SNSs have been used in decision-making [42–46], medical diagnosis [47], and so on.

In 2003, Kandasamy and Smarandache [48] defined the Neutrosophic Number (NN) by combining real numbers and indeterminate parts of the form $u + Iv$, where I is an indeterminate component, and u, v are real numbers. Several strategies for MAGDM have been proposed in NN environment [49–54]. Du et al. [55] combined SNS and NN to define the Simplified Neutrosophic Indeterminate Set (SNIS) and presented a new decision-making strategy. Köseoğlu et al. [56] extended SNS and defined the Simplified Neutrosophic Multiplicative Set (SNMS) and Simplified Neutrosophic Multiplicative Preference Relation (SNMPR).

Ye [57] defined the Single-Valued Neutrosophic Hesitant Fuzzy Set (SVNHFS) that encompasses the FS [1], IFS [7], hesitant FS [58], dual hesitant FS [59], and SVNS [10]. Theoretical developments and applications of SVNHFSs have been presented in several studies [60–67].

Neutrosophic Cubic Set (NCS) [68] was proposed by extending the cubic set [69]. Jun et al. [70] also proposed the NCS. Some theoretical developments and applications of NCSs have been made in MADM [71–74] and MAGDM [75, 76].

Bipolar NS (BNS) [77] was developed by extending bipolar FS [78] and NS [2]. BNSs have been utilized in dealing with MADM [79–83] and MAGDM [84].

Maji [85] grounded the Neutrosophic Soft Set (NSS) by combining Soft Set (SS) [86] and NS [2]. The NSSs have been utilized in MADM [87–92] and MAGDM [93–96]. Ali et al. [97] defined the bipolar neutrosophic SS by combining SS [86] and BNS [77].

Many investigators paid wide attention to the research of hybridizing Rough Set (RS) [98] and NS [2]. Broumi et al. [99, 100] defined the Rough Neutrosophic Set (RNS). Broumi and Smarandache [101] grounded the Interval RNS (IRNS). Yang et al. [102] defined the Single-Valued Neutrosophic RS (SVNRS) by exploring constructive and axiomatic strategies. Single-Valued Neutrosophic Multi-Granulation

RS (SVNMGRS) [103] was proposed by combining multi-granulation rough sets [104] with SVNS [10]. Jiao et al. [105] developed the three-way decision models utilizing Decision-Theoretic RS (DTRS) [106] with SVNS [10]. New theoretical developments and applications of the RNSs have been presented in the current neutrosophic literature [107–119]. Pramanik and Mondal [120] proposed the Rough BNS (RBNS). Recently, Pramanik [121] documented an overview of RNSs.

Yager [122] introduced the fuzzy bag or the Fuzzy Multi-Set (FMS) by extending the bag or multi-set [123, 124]. An element of a FMS can assume the same or different membership values more than once. In case of Intuitionistic Fuzzy Multi-Set (IFMS) [125], an element can assume membership and falsity values more than once. To overcome the shortcomings of FMS and IFMS, Smarandache [126] presented the n-valued refined neutrosophic logic. Smarandache [126] paved the way to define the Neutrosophic Refined (NR) set. Several investigators [127–136] studied the application of NR sets. Bao and Yang [137, 138] proposed the Single-Valued Neutrosophic Refined Rough Set (SVNRRS) by hybridizing NR sets with RSs.

Ye [139] extended the intuitionistic linguistic set [140] to Single-Valued Neutrosophic Linguistic (SVNL) set and SVNL number. Li et al. [141] defined the Neutrosophic Linguistic Set (NLS) and presented a comparison strategy for Neutrosophic Linguistic Numbers (NLNs). Ye [142] defined the Interval NLN (INLN) and employed it for MADM. Tian et al. [143] defined the Simplified NLS (SNLS) that combines the SNS and linguistic term set [144]. SNLS is capable of describing linguistic information to some extent. Tian et al. [143] employed the Simplified NLNs (SNLNs) for MADM.

Biswas et al. [145] and Ye [146] presented the Trapezoidal Neutrosophic Fuzzy Number (TrNFN) by extending the trapezoidal fuzzy numbers. TrNFNs have been utilized in MADM and MAGDM [147–153]. Biswas et al. [154] presented the ranking strategy for Single-Valued Neutrosophic Trapezoidal Number (SVNTrN) and employed it for MADM. Liang et al. [155] presented the Single-Valued Trapezoidal Neutrosophic Preference Relation (SVTrNPR) and the completely consistent SVTrNPR to solve MADM problems. Biswas et al. [156] extended the SVNTrN to interval trapezoidal neutrosophic number and employed it for MADM. Biswas et al. [157] defined the triangular fuzzy NSs and employed them for MADM. Abdel- Basset et al. [158] extended neutrosophic triangular number to propose the type-2 neutrosophic number. Chakraborty et al. [159] defined the neutrosophic pentagonal number and studied some of its properties. Karaaslan [160] presented the Gaussian Single-Valued Neutrosophic Number (SVNN) and employed it for MADM.

Researchers extended NSs to different sets such as bipolar neutrosophic refined sets [161], tri-complex rough neutrosophic set [162], rough neutrosophic hyper-complex set [163], quadripartitioned SVNS [164], plithogenic set [165, 166]. Smarandache [167, 168] further proposed the neutrosophic off/under/over sets by extending NSs.

NSs, SVNSs, and their hybrid extensions and applications can be found in the studies [121, 169–177].

Rest of the chapter is designed as follows: Sect. 2 presents the basics of the NSs. Section 3 presents the triangular fuzzy number NS. Section 4 presents the trapezoidal fuzzy NS. Section 5 describes single-valued pentagonal neutrosophic numbers. Section 6 describes the cylindrical neutrosophic single-valued sets. Section 7 describes the neutrosophic numbers. Section 8 describes some applications of NSs. Section 9 describes the extensions of the NSs. Section 10 presents the direction of new research. Section 11 presents conclusions.

2 Basics of Neutrosophic Sets

2.1 Neutrosophic Set

Let Q be a space of points with a generic element ω in Q. An NS [2] ϕ in Q is characterized by a truth. Membership Function (MF) ξ_ϕ, an indeterminacy MF ψ_ϕ, a falsity MF ζ_ϕ and is presented as:

$$\phi = \langle \omega, \xi_\phi(\omega), \psi_\phi(\omega), \zeta_\phi(\omega) > /, \omega \in \Omega \rangle.$$

Here, $\xi_\phi(\omega), \psi_\phi(\omega), \zeta_\phi(\omega)$ in Ω denote the subsets of $]^{-}0, 1^{+}[$ such that $\xi_\phi(\omega)$: $Q \rightarrow]^{-}0, 1^{+}[$, $\psi_\phi(\omega)$: $Q \rightarrow]^{-}0, 1^{+}[$, and $\zeta_\phi(\omega)$: $\Omega \rightarrow]^{-}0, 1^{+}[$.

Then,

$^{-}0 \leq \sup \xi_\phi(\omega) + \sup \psi_\phi(\omega) + \sup \zeta_\phi(\omega) \leq 3^{+}$.

2.2 Single-Valued Neutrosophic Set

An SVNS [10] χ in a universal set Ω is presented by a truth MF $\xi_\chi(\omega)$, an indeterminacy MF $\psi_\chi(\omega)$, and a falsity MF $\psi_\chi(\omega)$ such that.

$\xi_\chi(\omega), \psi_\chi(\omega), \zeta_\chi(\omega)$ are in [0, 1] for all $\omega \in \Omega$.

If Ω is continuous, χ is presented as

$$\chi = \int_\omega \langle \xi_\chi(\omega), \psi_\chi(\omega), \zeta_\chi(\omega) \rangle / \omega, \forall \omega \in \Omega.$$

If Ω is discrete, χ is presented as

$$\chi = \sum \langle \xi_\chi(\omega), \psi_\chi(\omega), \zeta_\chi(\omega) \rangle / \omega, \forall \omega \in \Omega$$

with $0 \leq \sup \xi_\chi(\omega) + \sup \psi_\chi(\omega) + \zeta_\chi(\omega) \leq 3, \omega \forall \in \Omega$.

An SVNS χ is also presented as.

$\chi = \langle \omega, \xi_\chi(\omega), \psi_\chi(\omega), \zeta_\chi(\omega) > /, \omega \in \Omega \rangle$, where $\xi_\chi(\omega), \psi_\xi(\omega), \zeta_\xi(\omega) \in [0, 1]$, for each ω in Ω. Therefore,

$$0 \leq \sup \xi_\chi(\omega) + \sup \psi_\chi(\omega) + \sup \zeta_\chi(\omega) \leq 3.$$

Note: Since SVNS is a subclass of NS, we use NS and SVNS equivalently throughout chapter.

For convenience, the triplet $(\xi_\chi(\omega), \psi_\chi(\omega), \zeta_\chi(\omega))$ is called as the SVNN and simply presented as $(\xi_\chi, \psi_\chi, \zeta_\chi)$.

2.2.1 Some Operations of SVNNs

Let $\eta_1 = (\alpha_1, \beta_1, \gamma_1)$ and $\eta_2 = (\alpha_2, \beta_2, \gamma_2)$ be any two SVNNs with $\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2 \in [0, 1]$, $(\alpha_1 + \beta_1 + \gamma_1) \in [0, 3]$ and $(\alpha_2 + \beta_2 + \gamma_2) \in [0, 3]$.

The following operations for SVNNs [171] hold

i. $\eta_1 \oplus \eta_2 = (\alpha_1 + \alpha_2 - \alpha_1\alpha_2, \beta_1, \beta_2, \gamma_1\gamma_2)$ [Summation]
ii. $\eta_1 \otimes \eta_2 = (\alpha_1\alpha_2, \beta_1 + \beta_2 - \beta_1\beta_2, \gamma_1 + \gamma_2 - \gamma_1\gamma_2)$ [Multiplication]
iii. $\lambda\eta_1 = (1 - (1 - \alpha_1)^\lambda, \beta_1^\lambda, \gamma_1^\lambda)$ [Scalar multiplication]
iv. $\eta_1^\lambda = (\alpha_1^\lambda, 1 - (1 - \beta_1)^\lambda, 1 - (1 - \gamma_1)^\lambda)), \lambda > 0.$

2.2.2 Score Function and Accuracy Function of SVNNs

Assume that $\eta_1 = (\alpha_1, \alpha_2, \alpha_3)$ is an SVNN. Score function denoted by $\Gamma(n_1)$, accuracy function denoted by $H(n_1)$ [178] of η_1 are, respectively, represented as.

i. $\Gamma(n_1) = \frac{(1 + \alpha_1 - 2\alpha_2 - \alpha_3)}{2}$, where $\Gamma(n_1) \in [-1, 1]$
ii. $H(n_1) = \alpha_1 - \alpha_2(1 - \alpha_1) - \alpha_3(1 - \alpha_2)$, where $H(n_1) \in [-1, 1]$
Nancy and Garg [179] presented the improved score function as follows:
iii. $S(n_1) = \frac{(1 + (\alpha_1 - 2\alpha_2 - \alpha_3)(2 - \alpha_1 - \alpha_3))}{2}$,

Clearly, if $\alpha_1 + \alpha_3 = 1$, $S(n_1)$ reduces to $\Gamma(n_1)$.

2.2.3 Comparison of SVNNs

Assume that $\chi_1 = (\alpha_1, \alpha_2, \alpha_3)$ and $\chi_2 = (\beta_1, \beta_2, \beta_3)$ be any two SVNNs. Comparison strategy [179] between χ_1 and χ_2 is presented as.

i. if $\Gamma(\chi_1) < \Gamma(\chi_2))$, then $\chi_1 \prec \chi_2$
ii. if $\Gamma(\chi_1) = \Gamma(\chi_2)$, then

- İf $S(\chi_1) < S(\chi_2)$, then $\chi_1 \prec \chi_2$
- $S(\chi_1) > S(\chi_2)$, then $\chi_1 \succ \chi_2$
- $S(\chi_1) = S(\chi_2)$, then $\chi_1 \approx \chi_2$.

2.3 Interval Neutrosophic Set

Let P be a space of points having generic element p in P.

An INS [25] φ in P is presented as $\varphi = \langle p, M_\varphi(p), N_\varphi(p), O_\varphi(p) > /, p \in P \rangle$.

Here, $M_\varphi(p) = [M^L_\varphi(p), M^U_\varphi(p)]$, $N_\varphi(p) = [N^L_\varphi(p), N^U_\varphi(p)]$, $O_\varphi(p) = [O^L_\varphi(p), O^U_\varphi(p)]$ and for each $p \in P$, $M_\varphi(p), N_\varphi(p), O_\varphi(p) \subseteq [0, 1]$.

For convenience, an Interval Neutrosophic Number (INN) η_1 is presented in the form: $\eta_1 = \langle [M^L_1, M^U_1], [N^L_1, N^U_1], [O^L_1, O^U_1] \rangle$

2.3.1 Operations on INNs

Let $\eta_1 = \langle [M^L_1, M^U_1], [N^L_1, N^U_1], [O^L_1, O^U_1] \rangle$ and $\eta_2 = \langle [M^L_2, M^U_2], [N^L_2, N^U_2], [O^L_2, O^U_2] \rangle$ be any two INNs. The operations for INNs [180] are presented as:

i. $\eta_1 \oplus \eta_2 = \langle [M^L_1 + M^L_2 - M^L_1 M^L_2, M^U_1 + M^U_2 - M^U_1 M^U_2], [N^L_1 N^L_2, N^U_1 N^U_2], [O^L_1 O^L_2, O^U_1 O^U_2] \rangle$

ii. $\eta_1 \otimes \eta_2 = \langle [M^L_1 M^L_2, M^U_1 M^U_2], [N^L_1 + N^L_2 - N^L_1 N^L_2, N^U_1 + N^U_2 - N^U_1 N^U_2], [O^L_1 + O^L_2 - O^L_1 O^L_2, O^U_1 + O^U_2 - O^U_1 O^U_2] \rangle$

iii. $\gamma\eta_1 = \langle [1 - (1 - M^L_1)^\gamma, 1 - (1 - M^U_1)^\gamma)], [(N^L_1)^\gamma, (N^U_1)^\gamma], [(O^L_1)^\gamma, (O^U_1)^\gamma] \rangle$

iv. $\eta_1^\gamma = \langle [(M^L_1)^\gamma, (M^U_1)^\gamma)], [1 - (1 - N^L_1)^\gamma, 1 - (1 - N^U_1)^\gamma)], [1 - (1 - O^L_1)^\gamma, 1 - (1 - O^U_1)^\gamma)] \rangle$ where $\gamma > 0$.

2.3.2 Score Function and Accuracy Functions of INNs

Assume that $\eta_1 = \langle [M^L_1, M^U_1], [N^L_1, N^U_1], [O^L_1, O^U_1] \rangle$ be an INN. The score function $\Gamma(\eta_1)$, accuracy function $H(\eta_1)$ [178] of η_1 are, respectively, presented as:

i. $\Gamma(\eta_1) = \left(\frac{1}{4}\right) \times [2 + M^L_1 + M^U_1 - 2N^L_1 - 2N^U_1 - O^L_1 - O^U_1]$, $\Gamma(\eta_1) \in [-1, 1]$

ii. $H(\eta_1) = \frac{M^L_1 + M^U_1 - N^U_1(1 - M^U_1) - N^L_1(1 - M^L_1) - O^U_1(1 - N^U_1) - \zeta^L_1(1 - N^L_1)}{2}$, $H(\eta_1) \in [-1, 1]$

2.3.3 Comparison INNs

The convenient strategy for comparing INNs [178] is described as follows:

Let $\eta_1 = \langle [M^L_1, M^U_1], [N^L_1, N^U_1], [O^L_1, O^U_1] \rangle$ and $\eta_2 = \langle [M^L_2, M^U_2], [N^L_2, N^U_2], [O^L_2, O^U_2] \rangle$ be any two INNs. Then,

i. If $\Gamma(\eta_1) \succ \Gamma(\eta_2)$, then $\eta_1 \succ \eta_2$
If $\Gamma(\eta_1) = \Gamma(\eta_2)$,and $H(\eta_1) \succ H(\eta_2)$, then $\eta_1 \succ \eta_2$.

2.4 Spherical Neutrosophic Set

Smarandache presented spherical NS [8], which is a generalization of spherical FS [181].

2.4.1 Single-Valued Spherical NS

A Single-Valued Spherical NS [8] of the universe of discourse Θ is presented as follows:

$$\chi = \langle \omega, \xi_\chi(\omega),\ \psi_\chi(\omega), \zeta_\chi(\omega) > /,\ \omega \in \Theta \rangle.$$

Here, $\forall w \in \Theta$, the functions $\xi_\chi(\omega),\ \psi_\chi(\omega), \zeta_\chi(\omega) : \Omega \to [0,\ \sqrt{3}]$ indicate the degrees of truth, indeterminacy, and falsity MF, respectively, and $0 \le \xi_\chi^2(\omega) + \psi_\chi^2(\omega) +\ \zeta_\chi^2(\omega) \le 3$.

Single-Valued Spherical Neutrosophic Number (SVSpNN)

Smarandache [182] presented the SVSpNN having the form: (q, r, s) where $q, r, s \in [0,\ \sqrt{3})$ and $q^2 + r^2 + s^2 \le 3$.

SVSrNN is the generalization of Single-Valued Pythagorean Fuzzy Number (SVPFN) having the form: (q, r) with $q,\ r \in [0,\ 2]$ and $q^2 + r^2 \le 2$.

Interval-Valued Spherical Neutrosophic Number (ISpNN)

Smarandache [182] presented the ISpNN that has the form:$(q,\ r, s)$ where the real intervals $q,\ r, s \subseteq [0,\ \sqrt{3}]$ and $q^2 + r^2 + s^2 \subseteq [0,\ 3]$.

2.4.2 n-Hyper Spherical Neutrosophic Set (n-HSpNS)

Single-valued n-HSpNS [8] is a generalization of the spherical NS in the universe of discourse Ω, for $n \ge 1$. It is defined as $\chi = \langle \omega, \xi_\chi(\omega),\ \psi_\chi(\omega), \zeta_\chi(\omega) > /,\ \omega \in \Omega \rangle$, where, $\forall \omega \in \Omega$, the functions $\xi_\chi(\omega),\ \psi_A(\omega), \zeta_A(\omega) : \Omega \to [0,\ \sqrt[n]{3}\,]$ represent the degrees of truth, indeterminacy, and falsity MF, respectively, and $0 \le \xi_\chi^n(\omega) +\ \psi_\chi^n(\omega) + \zeta_\chi^n(\omega) \le 3$.

3 Triangular Fuzzy Number Neutrosophic Set (TFNNS)

Biswas et al. [157] hybridized the Triangular Fuzzy Number (TFN) with SVNSs to define the TFNNS.

3.1 TFNNS

Let Ω be the finite universe of discourse and $\theta[0, 1]$ be the set of all TFNs on [0, 1]. A TFNNS φ in Ω is presented by

$$\varphi = \{\langle \omega, \xi_\varphi(\omega), \psi_\varphi(\omega), \zeta_\varphi(\omega)\ \rangle | \omega \in \Omega\},$$

where $\xi_\varphi(\omega) : \Omega \to \theta[0, 1]$,$\psi_\varphi(\omega) : \Omega \to \theta[0, 1]$, and $\zeta_\varphi(\omega) : \Omega \to \theta[0, 1]$.

The TFNs $\xi_{\varphi}(\omega) = \left(\xi^1_{\varphi}(\omega), \xi^2_{\varphi}(\omega),\ \xi^3_{\varphi}(\omega)\right)$,$\psi_{\varphi}(\omega) = \left(\psi^1_{\varphi}(\omega), \psi^2_{\varphi}(\omega),\right.$ $\left.\psi^3_{\varphi}(\omega)\right)$, and $\zeta_{\varphi}(\omega) = \left(\zeta^1_{\varphi}(\omega), \zeta^2_{\varphi}(\omega),\ \zeta^3_{\varphi}(\omega)\right)$ present the degree of truth, indeterminacy, and falsity MF, respectively, of $\omega \in \varphi$, $\forall \omega \in \Omega$ and $0 \le \xi_{\varphi}(\omega) + \psi_{\varphi}(\omega) + \zeta_{\varphi}(\omega) \le 3$.

3.2 Triangular Fuzzy Neutrosophic Number (TFNN)

For notational convenience, we consider $\phi = \langle(\alpha, \beta, \gamma), (\rho, \sigma, \tau), (l, m, n)\rangle$ as a TFNN [157] where $\left(\xi^1_{\varphi}(\omega), \xi^2_{\varphi}(\omega),\ \xi^3_{\varphi}(\omega)\right) = (\alpha, \beta, \gamma)$, $\left(\psi^1_{\varphi}(\omega), \psi^2_{\varphi}(\omega),\ \psi^3_{\varphi}(\omega)\right)$ $= (\rho, \sigma, \tau)$, and $\left(\zeta^1_{\varphi}(\omega), \zeta^2_{\varphi}(\omega),\ \zeta^3_{\varphi}(\omega)\right) = (l, m, n)$.

3.3 Operations on TFNNs

Let $\chi_1 = \langle(\alpha_1, \beta_1, \gamma_1), (\rho_1, \sigma_1, \tau_1), (l_1, m_1, n_1)\rangle$ and $\chi_2 = \langle(\alpha_2, \beta_2, \gamma_2), (\rho_2, \sigma_2, \tau_2), (l_2, m_2, n_2)\rangle$ be any two TFNNs in $\Re$. The basic operations for TFNNs [157] hold good:

i. $$\chi_1 \oplus \chi_2 = \langle(\alpha_1 + \alpha_2 - \alpha_1\alpha_2, \beta_1 + \beta_2 - \beta_1\beta_2,\ \gamma_1 + \gamma_2 - \gamma_1\gamma_2), (\rho_1\rho_2, \sigma_1\sigma_2, \tau_1\tau_2), (l_1 l_2, m_1 m_2, n_1 n_2)\rangle;$$

ii. $$\chi_1 \otimes \chi_2 = \left\langle \begin{array}{r} (\alpha_1\alpha_2, \beta_1\beta_2, \gamma_1\gamma_2), (\rho_1 + \rho_2 - \rho_1\rho_2, \sigma_1 + \sigma_2 - \sigma_1\sigma_2, \tau_1 + \tau_2 - \tau_1\tau_2), \\ (l_1 + l_2 - l_1 l_2, m_1 + m_2 - m_1 m_2, n_1 + n_2 - n_1 n_2) \end{array} \right\rangle;$$

iii. $$\lambda\chi_1 = \left\langle \left(1 - (1 - \varepsilon_1)^{\lambda}, 1 - (1 - \beta_1)^{\lambda}, 1 - (1 - \gamma_1)^{\lambda}\right), \left(\rho_1^{\lambda}, \sigma_1^{\lambda}, \tau_1^{\lambda}\right), \left(l_1^{\lambda}, m_1^{\lambda}, n_1^{\lambda}\right) \right\rangle$$
for $\lambda > 0$ and

iv. $$\chi_1^{\lambda} = \left\langle \begin{array}{r} \left(\alpha_1^{\lambda}, \beta_1^{\lambda}, \gamma_1^{\lambda}\right), \left(1 - (1 - \rho_1)^{\lambda}, 1 - (1 - \sigma_1)^{\lambda}, 1 - (1 - \tau_1)^{\lambda}\right), \\ \left(1 - (1 - l_1)^{\lambda}, 1 - (1 - m_1)^{\lambda}, 1 - (1 - n_1)^{\lambda}\right) \end{array} \right\rangle$$ for $\lambda > 0$.

The operations presented in Sect. 26.3 satisfy the following properties:

i. (Commutativity) : $\chi_1 \oplus \chi_2 = \chi_2 \oplus \chi_1;\ \chi_1 \otimes \chi_2 = \chi_2 \otimes \chi_1;$
ii. (Distributivity) : $\lambda(\chi_1 \oplus \chi_2) = \lambda\chi_1 \oplus \lambda\chi_2;\ (\chi_1 \otimes \chi_2)^{\lambda} = \chi_1^{\lambda} \otimes \chi_2^{\lambda}$, $\lambda > 0$, and
iii. (Associativity) : $\lambda_1\chi_1 \oplus \lambda_2\chi_1 = (\lambda_1 + \lambda_2)\chi_1;\ \chi_1^{\lambda_1} \oplus \chi_1^{\lambda_2} = \chi_1^{(\lambda_1 + \lambda_2)}$, $\lambda_1, \lambda_2 > 0$.

3.4 Score and Accuracy Function of TFNN

Let $\chi_1 = \langle(\alpha_1, \beta_1, \gamma_1), (\rho_1, \sigma_1, \tau_1), (l_1, m_1, n_1)\rangle$ be a TFNN in $\Re$. The score function [157] $\Gamma(\chi_1)$, the accuracy function [157] $H(\chi_1)$ of χ_1 are, respectively, presented as:

$$\Gamma(\chi_1) = \frac{1}{12}[8 + (\alpha_1 + 2\beta_1 + \gamma_1) - (\rho_1 + 2\sigma_1 + \tau_1) - (l_1 + 2m_1 + n_1)],$$
$$\Gamma(\chi_1) \in [1, 0]$$

$$H(\chi_1) = \frac{1}{4}[(\alpha_1 + 2\beta_1 + \gamma_1) - (l_1 + 2m_1 + n_1)],\ H(\chi_1) \in [-1, 1]$$

For $\chi^+ = \langle(1,1,1), (0,0,0), (0,0,0)\rangle$ and $\chi^- = \langle(0,0,0), (1,1,1), (1,1,1)\rangle$, $\Gamma(\chi^+)$ = 1 and $\Gamma(\chi^-) = 0$.

For $\chi^+ = \langle(1,1,1), (0,0,0), (0,0,0)\rangle$ and $\chi^- = \langle(0,0,0), (1,1,1), (1,1,1)\rangle$, $H(\chi^+)$ = 1 and $H(\chi^-) = 0$.

3.5 Comparison of TFNNs

Assume that $\chi_1 = \langle(\alpha_1, \beta_1, \gamma_1), (\rho_1, \sigma_1, \tau_1), (l_1, m_1, n_1)\rangle$ and $\chi_2 = \langle(\alpha_2, \beta_2, \gamma_2), (\rho_2, \sigma_2, \tau_2), (l_2, m_2, n_2)\rangle$ are any two TFNNs in $\Re$. If $\Gamma(\chi_i)$ and $H(\chi_i)$ denote, respectively, the score and accuracy function of $\chi_i (i = 1, 2)$, then the ranking of TFNNs [157] is presented as:

i. If $\Gamma(\chi_1) > \Gamma(\chi_2)$, then χ_1 is greater than χ_2 that is $\chi_1 \succ \chi_2$;
ii. If $\Gamma(\chi_1) = \Gamma(\chi_2)$ and $H(\chi_1) > H(\chi_2)$, then χ_1 is greater than χ_2, that is,$\chi_1 \succ \chi_2$;
iii. If $\Gamma(\chi_1) = \Gamma(\chi_2)$,$\Gamma(\chi_1) = \Gamma(\chi_2)$, then χ_1 is indifferent to χ_2, that is, $\chi_1 \approx \chi_2$.

3.6 Triangular Fuzzy Neutrosophic Number Arithmetic Averaging (TFNNAA) Operator

3.6.1 Triangular Fuzzy Neutrosophic Number Weighted Arithmetic Averaging (TFNNWAA) Operator

Let $\theta_\imath = \langle(c_\imath, d_\imath, e_\imath), (f_\imath, g_\imath, h_\imath), (o_\imath, p_\imath, q_\imath)\rangle (\imath = 1, 2, \ldots, r)$ be a collection TFNNs in $\Re$ and let TFNNWAA : $\Sigma^r \to \Sigma$.

If $\mathrm{TFNNWAA}(\theta_1, \theta_2, \ldots, \theta_r) = \omega_1\theta_1 \oplus \omega_2\theta_2 \oplus \cdots \oplus \omega_r\theta_r = \overset{r}{\underset{k=1}{\oplus}} (\omega_k \theta_k)$,

then the function $\mathrm{TFNNWAA}(\theta_1, \theta_2, \ldots, \theta_r)$ is called the TFNNWAA operator, where the weight of $\theta_\imath (\imath = 1, 2, \ldots, r)$ is denoted by $\omega_\imath \in [0, 1]$ and $\sum_{\imath=1}^{r} \omega_\imath = 1$.

If $\omega = (1/r, 1/r, ..., 1/r)^T$, then the TFNNWAA$(\theta_1, \theta_2, ..., \theta_r)$ operator reduces to TFNNAA operator:

TFNNAA$(\theta_1, \theta_2, ..., \theta_r) = \frac{1}{r}(\theta_1 \oplus \theta_2 \oplus \cdots \oplus \theta_r)$

Theorem 3.6.1 *Let* $\theta_\imath = \langle (c_\imath, d_\imath, e_\imath), (f_\imath, g_\imath, h_\imath), (o_\imath, p_\imath, q_\imath)\rangle (\imath = 1, 2, ..., r)$ *be a collection of* TFNNs *in* $\Re$. *Then*, TFNNWAA$_\omega(\theta_1, \theta_2, ..., \theta_r) = \omega_1\theta_1 \oplus \omega_2\theta_2 \oplus \cdots \oplus \omega_r\theta_r) = \overset{r}{\underset{\imath=1}{\oplus}}(\omega_\imath\theta_\imath)$.

$$= \left\langle \begin{array}{l} \left(1 - \prod_{\imath=1}^{r}(1-c_\imath)^{\omega_\imath}, 1 - \prod_{\imath=1}^{r}(1-d_\imath)^{\omega_\imath}, 1 - \prod_{\imath=1}^{r}(1-e_\imath)^{\omega_\imath}\right), \\ \left(\prod_{\imath=1}^{r} f_\imath^{\omega_\imath}, \prod_{\imath=1}^{r} g_\imath^{\omega_\imath}, \prod_{\imath=1}^{r} h_\imath^{\omega_\imath}\right), \left(\prod_{\imath=1}^{r} o_\imath^{\omega_\imath}, \prod_{\imath=1}^{r} p_\imath^{\omega_\imath}, \prod_{\imath=1}^{r} q_\imath^{\omega_\imath}\right) \end{array} \right\rangle, \text{ where } \omega_\imath \in$$

$[0, 1]$ *denotes the weight of* $\theta_\imath(\imath = 1, 2, ..., r)$ and $\sum_{\imath=1}^{r} \omega_\imath = 1$.

Proof For proof, see Biswas et al. [157].

3.6.2 Triangular Fuzzy Number Neutrosophic Geometric Averaging (TFNNGA) Operator

Triangular Fuzzy Number Neutrosophic Weighted Geometric Averaging (TFNNWGA) Operator

Let $\theta_\imath = \langle (c_\imath, d_\imath, e_\imath), (f_\imath, g_\imath, h_\imath), (o_\imath, p_\imath, q_\imath)\rangle$ $(\imath = 1, 2, ..., r)$ be a collection of TFNNs in $\Re$ and TFNNWGA : $\Pi^r \to \Pi$.

If

TFNNWGA$_\omega(\theta_1, \theta_2, ..., \theta_r) = \theta_1^{\omega_1} \otimes \theta_2^{\omega_2} \otimes \cdots \otimes \theta_r^{\omega_r} = \overset{r}{\underset{\imath=1}{\otimes}}\left(\theta_\imath^{\omega_\imath}\right)$

then TFNNWGA$_\omega(\theta_1, \theta_2, ..., \theta_r)$ is called the TFNNWGA operator where $\omega_\imath \in [0, 1]$ is the exponential weight of $\theta_k(\imath = 1, 2, ..., r)$ such that $\sum_{i=1}^{r} w_k = 1$.

If $\omega = (1/r, r, ..., 1/r)^T$, then the TFNNWGA$(\theta_1, \theta_2, ..., \theta_r)$ operator reduces to TFNNGA operator:

TFNNGA$(\theta_1, \theta_2, ..., \theta_r) = (\theta_1 \otimes \theta_2 \otimes \cdots \otimes \theta_r)^{\frac{1}{r}}$.

Theorem 3.6.2 *Let* $\theta_\imath = \langle (c_\imath, d_\imath, e_\imath), (f_\imath, g_\imath, h_\imath), (o_\imath, p_\imath, q_\imath)\rangle (\imath = 1, 2, ..., r)$ *is a collection of TFNNs in* $\Re$. *Then*,TFNNWGA$_\omega(\theta_1, \theta_2, ..., \theta_r) = \theta_1^{\omega_1} \otimes \theta_2^{\omega_2} \otimes \cdots \otimes \theta_n^{\omega_\imath} = \overset{r}{\underset{\imath=1}{\otimes}}\left(\theta_\imath^{\omega_\imath}\right)$

$$\left(\prod_{\iota=1}^{r} c_\iota^{\omega_\iota}, \prod_{\iota=1}^{r} d_\iota^{\omega_\iota}, \prod_{\iota=1}^{r} e_\iota^{\omega_\iota}\right),$$

$$= \left\langle \begin{array}{l} \left(1-\prod_{\iota=1}^{r}(1-f_\iota)^{\omega_\iota}, 1-\prod_{\iota=1}^{r}(1-g_\iota)^{\omega_\iota}, 1-\prod_{\iota=1}^{r}(1-h_\iota)^{\omega_\iota}\right), \\ \left(1-\prod_{\iota=1}^{r}(1-o_\iota)^{\omega_\iota}, 1-\prod_{\iota=1}^{r}(1-p_\iota)^{\omega_\iota}, 1-\prod_{\iota=1}^{r}(1-q_\iota)^{\omega_\iota}\right) \end{array} \right\rangle \text{ is a TFNN,}$$

where $\omega_k \in [0, 1]$ *denotes the weight vector of* TFNN $\theta(\iota = 1, 2, ..., r)$ *such that* $\sum_{\iota=1}^{r} \omega_\iota = 1$.

Proof For proof, see Biswas et al. [157].

4 Trapezoidal Fuzzy Number NS

Ye [146] and Biswas et al. [145] combined Trapezoidal Fuzzy Number (TrFN) with SVNS to define Trapezoidal Fuzzy Number NS (TrFNNS).

4.1 TrFNNS

Assume that Θ is the finite universe of discourse. A TrFNNS θ in Θ is presented as:

$\theta = \{\langle x, o_\theta(x), p_\theta(x), q_\theta(x)\rangle | x \in \Theta\}$, where $O_\theta(x) \subset [0, 1]$,$p_\theta(x) \subset [0, 1]$ $q_\theta(x) \subset [0, 1]$ are trpezoidal fuzzy numbers and $o_\theta(x) = \left(o^1_\theta(x), o^2_\theta(x), o^3_\theta(x), o^4_\theta(x)\right) : \Theta \rightarrow [0, 1]$,$p_\theta(x) = \left(p^1_\theta(x), p^2_\theta(x), p^3_\theta(x), p^4_\theta(x)\right) : \Theta \rightarrow [0, 1]$, and $q_\theta(x) = \left(q^1_\theta(x), q^2_\theta(x), q^3_\theta(x), q^4_\theta(x)\right) : \Theta \rightarrow [0, 1]$ present, respectively, the degrees of truth, indeterminacy, and falsity MF of x in θ and for every $x \in \Theta$ and $0 \le o^4_\theta(x) + p^4_\theta(x) + q^4_\theta(x) \le 3$.

4.2 Trapezoidal Fuzzy Neutrosophic Number (TrFNN)

A TrFNN [145] χ_φ is presented as:

$\chi_\varphi = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4), (\beta_1, \beta_2, \beta_3, \beta_4), (\gamma_1, \gamma_2, \gamma_3, \gamma_4)\rangle$ in a universe of discourse W with $\alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4$, $\beta_1 \le \beta_2 \le \beta_3 \le \beta_4$ and $\gamma_1 \le \gamma_2 \le \gamma_3 \le \gamma_4$. Here,$\chi_\varphi$ is defined as follows:

Its truth MF is presented as:

$$\xi_{\chi_\varphi}(\omega) = \begin{cases} \frac{\omega-\alpha_1}{\alpha_2-\alpha_1}, & \alpha_1 \le \omega \le \alpha_2 \\ 1, & \alpha_2 \le \omega \le \alpha_3 \\ \frac{\alpha_4-\omega}{\alpha_4-\alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise}. \end{cases}$$

Its indeterminacy MF is presented as:

$$\psi_{\chi_\varphi}(\omega) = \begin{cases} \frac{\omega-\beta_1}{\beta_2-\beta_1}, & \beta_1 \le \omega \le \beta_2 \\ 1, & \beta_2 \le \omega \le \beta_3 \\ \frac{\beta_4-\omega}{\beta_4-\beta_3}, & \beta_3 \le \omega \le \beta_4 \\ 0, & \text{otherwise}. \end{cases}$$

and its falsity MF is presented as:

$$\zeta_{\chi_\varphi}(\omega) = \begin{cases} \dfrac{\omega-\gamma_1}{\gamma_2-\gamma_1}, & \gamma_1 \le \omega \le \gamma_2 \\ 1, & \gamma_2 \le \omega \le \gamma_3 \\ \dfrac{\gamma_4-\omega}{\gamma_4-\gamma_3}, & \gamma_3 \le \omega \le \gamma_4 \\ 0 & \text{otherwise}. \end{cases}$$

Let $\chi_1 = \langle(\alpha_1, \beta_1, \gamma_1, \delta_1), (\rho_1, \sigma_1, \tau_1, \upsilon_1), (l_1, m_1, n_1, o_1)\rangle$ and $\chi_2 = \langle(\alpha_2, \beta_2, \gamma_2, \delta_2), (\rho_2, \sigma_2, \tau_2, \upsilon_2), (l_2, m_2, n_2, 0_2)\rangle$. be any two TrFNNs in $\Re$. Then, the operational rules for χ_1 and χ_2 are presented as.

i. $$\chi_1 \oplus \chi_2 = \left\langle \begin{array}{r} (\alpha_1+\alpha_2-\alpha_1\alpha_2, \beta_1+\beta_2-\beta_1\beta_2, \gamma_1+\gamma_2-\gamma_1\gamma_2, \delta_1+\delta_2-\delta_1\delta_2), \\ (\rho_1\rho_2+\sigma_1\sigma_2+\tau_1\tau_2+\upsilon_1\upsilon_2), (l_1 l_2+m_1 m_2+n_1 n_2+o_1 o_2) \end{array} \right\rangle$$

ii. $$\chi_1 \otimes \chi_2 = \left\langle \begin{array}{r} (\alpha_1\alpha_2, \beta_1\beta_2, \gamma_1\gamma_2, \delta_1\delta_2), \\ (\rho_1+\rho_2-\rho_1\rho_2, \sigma_1+\sigma_2-\sigma_1\sigma_2, \tau_1+\tau_2-\tau_1\tau_2, \upsilon_1+\upsilon_2-\upsilon_1\upsilon_2), \\ (l_1+l_2-l_1 l_2, m_1+m_2-m_1 m_2, n_1+n_2-n_1 n_2, o_1+o_2-o_1 o_2) \end{array} \right\rangle;$$

iii. $$\eta\chi_1 = \left\langle \begin{array}{c} \begin{pmatrix} 1-(1-\alpha_1)^\eta, 1-(1-\beta_1)^\eta, \\ 1-(1-\gamma_1)^\eta, 1-(1-\delta_1)^\eta \end{pmatrix}, \\ (\rho_1^\eta, \sigma_1^\eta, \tau_1^\eta, \upsilon_1^\eta), \ (l_1^\eta, m_1^\eta, n_1^\eta, o_1^\eta) \end{array} \right\rangle \text{ for } \eta > 0$$

iv. $$(\chi_1)^\eta = \left\langle \begin{array}{r} (\alpha_1^\eta, \beta_1^\eta, \gamma_1^\eta, \delta_1^\eta), \\ (1-(1-\rho_1)^\eta, 1-(1-\sigma_1)^\eta, 1-(1-\tau_1)^\eta, 1-(1-\upsilon_1)^\eta), \\ (1-(1-l_1)^\eta, 1-(1-m_1)^\eta, 1-(1-n_1)^\eta, 1-(1-o_1)^\eta), \end{array} \right\rangle \text{ for } \eta > 0.$$

v. $\chi_1 = \chi_2$, if $(\alpha_1, \beta_1, \gamma_1, \delta_1) = (\alpha_2, \beta_2, \gamma_2, \delta_2)$; $(\rho_1, \sigma_1, \tau_1, \upsilon_1) = (\rho_2, \sigma_2, \tau_2, \upsilon_2)$; $(l_1, m_1, n_1, o_1) = (l_2, m_2, n_2, 0_2)$

4.3 Expected Value (EV) and Expected Interval (EI) of TrFNN

The EI and the EV [145] of the truth MF $\xi_{\chi_1}(\omega) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \begin{cases} \frac{\omega-\alpha_1}{\alpha_2-\alpha_1}, & \alpha_1 \le \omega \le \alpha_2 \\ 1, & \alpha_2 \le \omega \le \alpha_3 \\ \frac{\alpha_4-\omega}{\alpha_4-\alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise}. \end{cases}$ of χ_1 in a universe of discourse W are defined as follows:

$$\mathrm{EI}\xi_{\chi_1}(\omega) = \left[\frac{(\alpha_1+\alpha_2)}{2}, \frac{(\alpha_3+\alpha_4)}{2}\right]$$

$$\mathrm{EV}\xi_{\chi_1}(\omega) = \frac{(\alpha_1+\alpha_2+\alpha_3+\alpha_4)}{4}$$

In similar way, the EI and the EV of the indeterminacy MF $\psi_{\chi_1}(\omega) = \begin{cases} \frac{\omega-\beta_1}{\beta_2-\beta_1}, & \beta_1 \le \omega \le \beta_2 \\ 1, & \beta_2 \le \omega \le \beta_3 \\ \frac{\beta_4-\omega}{\beta_4-\beta_3}, & \beta_3 \le \omega \le \beta_4 \\ 0, & \text{otherwise}. \end{cases}$
of χ_1 are defined as:

$$\mathrm{EI}\psi_{\chi_1}(\omega) = \left[\frac{(\beta_1+\beta_2)}{2}, \frac{(\beta_3+\beta_4)}{2}\right]$$

$$\mathrm{EV}\psi_{\chi_1}(\omega) = \frac{(\beta_1+\beta_2+\beta_3+\beta_4)}{4}$$

and the EI and the EV of the falsity MF

$$\zeta_{\chi_1}(\omega) = \begin{cases} \dfrac{\omega-\gamma_1}{\gamma_2-\gamma_1}, & \gamma_1 \le \omega \le \gamma_2 \\ 1, & \gamma_2 \le \omega \le \gamma_3 \\ \dfrac{\gamma_4-\omega}{\gamma_4-\gamma_3}, & \gamma_3 \le \omega \le \gamma_4 \\ 0 & \text{otherwise}. \end{cases}$$

of χ_1 are defined as follows:

$$\mathrm{EI}(\zeta_{\chi_1}(\omega)) = \left[\frac{(\gamma_1+\gamma_2)}{2}, \frac{(\gamma_3+\gamma_4)}{2}\right]$$

$$\mathrm{EV}(\zeta_1(\omega)) = \frac{(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4)}{4}$$

4.4 Truth Favourite Relative Expected Value (TrFREV) of TrFNN

Let $a = \langle(\alpha_1, \beta_1, \gamma_1, \delta_1), (\rho_1, \sigma_1, \tau_1, \upsilon_1), (l_1, m_1, n_1, o_1)\rangle$ be a TrFNN in $\Re$. Then TrFREV [145] of a is defined as:

$$\mathrm{EV}^{\mathrm{truth}}(a) = \frac{3EV(\xi_a(\omega))}{\mathrm{EV}(\xi_a(\omega)) + \mathrm{EV}(\psi_a(\omega)) + \mathrm{EV}(\zeta_a(\omega))},$$

where $\mathrm{EV}(\xi_a(w))$, $\mathrm{EV}(\psi_a(\omega))$, and $\mathrm{EV}(\zeta_a(\omega))$ denote, respectively, the EVs of truth, indeterminacy, and falsity MF of a.

4.5 Expected Value Theorem

Assume that $\phi_1 = \langle(\alpha_1, \beta_1, \gamma_1, \delta_1), (\rho_1, \sigma_1, \tau_1, \upsilon_1), (l_1, m_1, n_1, o_1)\rangle$ is a TrFNN in $\Re$, then the TFREV of ϕ_1 is

$$\mathrm{EV}^{\mathrm{truth}}(a) = \frac{3\sum_{i=1}^{4}\alpha_i}{\left(\sum_{i=1}^{4}(\alpha_i + \beta_i + \gamma_i)\right)}$$

Proof For proof, see [145].

4.6 Single-Valued Trapezoidal Neutrosophic Number (SVTrNN)

An SVTrNN [146] (SVTNN) $\alpha = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha)\rangle$ is a special NS on the real number set R, whose truth MF, indeterminacy MF, and a falsity MF are presented as:

$$\xi_\alpha(\omega) = \begin{cases} (\omega - \alpha_1)\xi_\alpha/(\alpha_2 - \alpha_1), & (\alpha_1 \le \omega \le \alpha_2) \\ \xi_\alpha, & (\alpha_2 \le \omega \le \alpha_3) \\ (\alpha_4 - \omega)\xi_\alpha/(\alpha_4 - \alpha_3), & (\alpha_3 \le \omega \le \alpha_4) \\ 0, & \text{otherwise} \end{cases}$$

$$\psi_\alpha(\omega) = \begin{cases} (\alpha_2 - \omega + \psi_a(\omega - \alpha_1))/(\alpha_2 - \alpha_1), & (\alpha_1 \le \omega \le \alpha_2) \\ \psi_\alpha, & (\alpha_2 \le \omega \le \alpha_3) \\ (\omega - \alpha_3 + \psi_\alpha(\alpha_4 - \omega))/(\alpha_4 - \alpha_3), & (\alpha_3 \le x \le \alpha_4) \\ 1, & \text{otherwise} \end{cases}$$

$$\zeta_\alpha(\omega) = \begin{cases} (\alpha_2 - \omega + \zeta_a(\omega - \alpha_1))/(\alpha_2 - \alpha_1), & (\alpha_1 \le \omega \le \alpha_2) \\ \zeta_\alpha, & (\alpha_2 \le \omega \le \alpha_3) \\ (\omega - \alpha_3 + \zeta_\alpha(\alpha_4 - \omega))/(\alpha_4 - \alpha_3), & (\alpha_3 \le x \le \alpha_4) \\ 1, & \text{otherwise} \end{cases}$$

where $0 \le \xi_\alpha \le 1$, $0 \le \psi_\alpha \le 1$, $0 \le \zeta_\alpha \le 1$; and $0 \le \xi_\alpha + \psi_\alpha + \zeta_\alpha \le 3$; $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \Re$.

Here, $\alpha = <(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha) >$ is called a positive (+ve) SVTrNN, if $\alpha_1 > 0$.

Similarly, $\alpha = <(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha) >$ reduces to a negative (–ve) SVTrNN, if $\alpha_4 \le 0$,

When $0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4 \le 1)$, and $0 \le \xi_\alpha \le 1$, $0 \le \psi_\alpha \le 1$, $0 \le \zeta_\alpha \le 1$, α is called a normalized SVTrNN.

4.6.1 Score Function of SVTrNNs

Let $\alpha = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha)\rangle$ be an SVTrNN. Then, the score and accuracy function [183] of α are denoted by $\Gamma(\alpha)$ and $H(\alpha)$, respectively, and are presented as:

i. $\Gamma(\alpha) = \left(\frac{1}{12}\right)(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) \times (2 + \xi_\alpha - \psi_\alpha - \zeta_\alpha)$
ii. $H(\alpha) = \left(\frac{1}{12}\right)(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) \times (2 + \xi_\alpha - \psi_\alpha + \zeta_\alpha)$

4.6.2 Ranking of SVTrNN

Let $\alpha = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha)\rangle$ and $\beta = \langle(\beta_1, \beta_2, \beta_3, \beta_4); (\xi_\beta, \psi_\beta, \zeta_\beta)\rangle$ be any two SVTrNNs. Ranking of SVTrNNs [183] is presented as follows:

(i) When $\Gamma(\alpha) < \Gamma(\beta)$, then $\alpha < \beta$
(ii) When $\Gamma(\alpha) = \Gamma(\beta)$ and if $H(\alpha) < H(\beta)$, then $\alpha < \beta$
(iii) When $\Gamma(\alpha) = \Gamma(\beta)$ and $H(\alpha) = H(\beta)$, then $\alpha = \beta$.

4.6.3 Centre of Gravity

Let $\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ be a TrFN on $\Re$, and $\alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4$; then, the centre of gravity (COG) of α [155, 184, 185] defined by

$$\mathrm{COG}(\alpha) = \left\{ \begin{array}{ll} \alpha_1, & \text{if} \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \\ \frac{1}{3}\left[\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{\alpha_4 - \alpha_3 - \alpha_{21} - \alpha_1}{\alpha_4 + \alpha_3 - \alpha_2 - \alpha_1}\right], & \text{otherwise} \end{array} \right\}$$

4.6.4 Score function SVTrNN

Assume that $\alpha = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha)\rangle$ is an SVTrNN. Then, the score, accuracy, and certainty function [155] of α are denoted by $\Gamma(\alpha)$, $H(\alpha)$, and $\kappa(\alpha)$), respectively, and presented as:

$$\Gamma(\alpha) = COG(\alpha) \times \frac{(2 + \xi_\alpha - \psi_\alpha - \zeta_\alpha)}{3}$$
$$H(\alpha) = COG(\alpha) \times (\xi_\alpha - \zeta_\alpha)$$
$$\kappa(\alpha) = COG(\alpha) \times \xi_\alpha$$

4.6.5 Comparison of SVTrNNs

Assume that $\alpha = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4); (\xi_\alpha, \psi_\alpha, \zeta_\alpha)\rangle$, and $\beta = \langle(\beta_1, \beta_2, \beta_3, \beta_4);\ (\xi_\beta, \psi_\beta, \zeta_\beta)\rangle$ are any two SVTRNNs. Comparison between SVTrNNs [155] is presented as follows:

(i) When $\Gamma(\alpha) > \Gamma(\beta)$, then $\alpha > \beta$
(ii) When $\Gamma(\alpha) = \Gamma(\beta)$ and if $H(\alpha) > H(\beta)$, then $\alpha > \beta$
(iii) When $\Gamma(\alpha) = \Gamma(\beta)$ and $H(\alpha) < H(\beta)$, then $\alpha < \beta$
(iv) When $\Gamma(\alpha) = \Gamma(\beta)$, $H(\alpha) = H(\beta)$, and

- $\kappa(\alpha) > \kappa(\beta)$, then $\alpha > \beta$
- $\kappa(\alpha) < \kappa(\beta)$, then $\alpha < \beta$
- $\kappa(\alpha) = \kappa(\beta)$, then $\alpha = \beta$.

4.7 Interval TrNN (ITrNN)

Suppose that χ is an SVTrNN [156]. Its truth MF, indeterminacy MF, and falsity MF are defined as

$$\xi_\chi(\omega) = \begin{cases} \dfrac{(\omega - \alpha_1)a'_\chi}{(\alpha_2 - \alpha_1)}, & \alpha_1 \le \omega \le \alpha_2 \\ a'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{(\alpha_4 - \omega)a'_\chi}{(\alpha_4 - \alpha_3)}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

$$\psi_\chi(\omega) = \begin{cases} \dfrac{(\alpha_2 - \omega) + (\omega - \alpha_1)b'_\chi}{(\alpha_2 - \alpha_1)}, & \alpha_1 \le \omega \le \alpha_2 \\ b'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{\omega - \alpha_3 + (\alpha_4 - \omega)b'_\chi}{\alpha_4 - \alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

$$\zeta_\chi(\omega) = \begin{cases} \dfrac{\alpha_2 - \omega + (\omega - \alpha_1)c'_\chi}{\alpha_2 - \alpha_1}, & \alpha_1 \le \omega \le \alpha_2 \\ c'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{\omega - \alpha_3 + (\alpha_4 - \omega)c'_\chi}{\alpha_4 - \alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

where $0 \le \xi_\chi(\omega) \le 1,\ 0 \le \psi_\chi(\omega) \le 1, 0 \le \zeta_\chi(\omega) \le 1,\quad 0 \le \xi_\chi(\omega) + \psi_\chi(\omega) + \zeta_\chi(\omega) \le 3,\ \alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4 \in \Re$ and $a'_\chi, b'_\chi, c'_\chi \in [0,\ 1]$. Then, neutrosophic trapezoidal number χ is presented as $\chi = (\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4; a'_\chi, b'_\chi, c'_\chi)$.

4.7.1 Definition of ITrNN

Assume that $a'_\chi = [a'^l_\chi, a'^u_\chi], b'_\chi = [b'^l_\chi, b'^u_\chi], c'_\chi = [c'^l_\chi, c'^u_\chi,]$. Then, an ITrNN [156] χ denoted by $\chi = ([\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4]; a'_\chi, b'_\chi, c'_\chi)$ is defined as

$$\xi_\chi(\omega) = \begin{cases} \dfrac{(\omega - \alpha_1)a'_\chi}{(\alpha_2 - \alpha_1)}, & \alpha_1 \le \omega \le \alpha_2 \\ a'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{(\alpha_4 - \omega)a'_\chi}{(\alpha_4 - \alpha_3)}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

$$\psi_\chi(\omega) = \begin{cases} \dfrac{(\alpha_2 - \omega) + (\omega - \alpha_1)b'_\chi}{(\alpha_2 - \alpha_1)}, & \alpha_1 \le \omega \le \alpha_2 \\ b'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{\omega - \alpha_3 + (\alpha_4 - \omega)b'_\chi}{\alpha_4 - \alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

$$\zeta_\chi(\omega) = \begin{cases} \dfrac{\alpha_2 - \omega + (\omega - \alpha_1)c'_\chi}{\alpha_2 - \alpha_1}, & \alpha_1 \le \omega \le \alpha_2 \\ c'_\chi, & \alpha_2 \le \omega \le \alpha_3 \\ \dfrac{\omega - \alpha_3 + (\alpha_4 - \omega)c'_\chi}{\alpha_4 - \alpha_3}, & \alpha_3 \le \omega \le \alpha_4 \\ 0, & \text{otherwise.} \end{cases}$$

where $a'_\chi, b'_\chi, c'_\chi \subset [0, 1]$ denote interval numbers, $0 \le \sup(a'_\chi) + \sup(b'_\chi) + \sup(c'_\chi) \le 3$ and $\chi = ([\alpha_1, \alpha_2, \alpha_3, \alpha_4]; [a'^l_\chi, a'^u_\chi], [b'^l_\chi, b'^u_\chi], [c'^l_\chi, c'^u_\chi,]$ is said to be positive ITrNN if $\chi > 0$ and one of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ is not equal to zero.

4.7.2 Operations on ITrNNs

Let $\chi_1 = \left([\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}]; [p_1^l, p_1^u], [q_1^l, q_1^u], [r_1^l, r_1^u]\right)$ and $\chi_2 = ([\alpha_{21}, \alpha_{22}, \alpha_{23}, \alpha_{24}]; [p_2^l, p_2^u], [q_2^l, q_2^u], [r_2^l, r_2^u])$ be any two ITrNNs and $\lambda > 0$. Then, the following operations hold good [156].

i. $$\chi_1 \oplus \chi_2 == ([\alpha_{11} + \alpha_{21}, \alpha_{12} + \alpha_{22}, \alpha_{13} + \alpha_{23}, \alpha_{14} + \alpha_{24}]; [p_1^l + p_2^l - p_1^l p_2^l, p_1^u + p_2^u - p_1^u p_2^u], [q_1^l q_2^l, q_1^u q_2^u], [r_1^l r_2^l, r_1^u r_2^u])$$

$$\begin{aligned} \chi_1 \otimes \chi_2 = ([\alpha_{11}\,\alpha_{21}, \alpha_{12}\alpha_{22}, \alpha_{13}\,\alpha_{31}, \alpha_{14}\alpha_{24}]; [p_1^l p_2^l,\ p_1^u p_2^u], \\ \left[q_1^l + q_2^l - q_1^l q_2^l, {}_1^u + q_2^u - q_1^u q_2^u\right], \\ \left[r_1^l + r_2^l - r_1^l r_2^l,\ r_1^u + r_2^u - r_1^u r_2^u\right]) \end{aligned}$$

$$\lambda\chi_1 = ([\lambda\alpha_{11}, \lambda\,\alpha_{12}, \lambda\alpha_{13}, \lambda\alpha_{14}];$$
$$[1 - (1 - p_1^l)^\lambda, 1 - (1 - p_2^u)^\lambda], [(q_1^l)^\lambda, (q_2^u)^\lambda], [(r_1^l)^\lambda, (r_2^\lambda)^\lambda])$$

$$(\chi_1)^\lambda = (\left[(\alpha_{11})^\lambda, (\alpha_{12})^\lambda, (\alpha_{13})^\alpha, (\alpha_{14})^\alpha\right];$$
$$[(p_1^l)^\lambda, (p_1^u)^\lambda], [1 - (1 - q_1^l)^\lambda, 1 - (1 - q_1^u)^\lambda], [1 - (1 - r_1^l)^\lambda, 1 - (1 - r_1^u)^\lambda])$$

5 A Single-Valued Pentagonal Neutrosophic Number (SVPNN)

An SVPNN [159] χ is defined as

$$\chi = \left\langle \omega, \xi_\chi(\omega),\ \psi_A(\omega), \zeta_A(\omega) > /,\ \omega \in \Omega \right\rangle.$$

The truth MF $\xi_\chi(\omega) : \Re \to [0, \alpha]$, the indeterminacy MF $\psi_\chi(\omega) : \Re \to [\beta, 1]$, and the falsity MF $\zeta_\chi(\omega) : \Re \to [\gamma, 1]$ are presented as:

$$\xi_{\chi}(\omega)=\begin{cases}\xi_{\chi_{L1}}(\omega),\ p_1'\le\omega\le q_1'\\ \xi_{\chi_{L2}}(\omega),q_1'\le\omega\le r_1'\\ \tau,\quad \omega=r_1'\\ \xi_{\chi_{U1}}(\omega),\ r_1'\le\omega\le s_1'\\ \xi_{\chi_{U2}}(\omega),\ s_1'\le\omega\le t_1'\\ 0,\quad \text{otherwise}\end{cases},$$

$$\psi_{\chi}(\omega)=\begin{cases}\psi_{\chi_{L1}}(\omega),\ p_2'\le\omega\le q_2'\\ \psi_{\chi_{L2}}(\omega),q_2'\le\omega\le r_2'\\ \iota,\quad \omega=r_2'\\ \psi_{\chi_{U1}}(\omega),\ r_2'\le\omega\le s_2'\\ \psi_{\chi_{U2}}(\omega),\ s_2'\le\omega\le t_2'\\ 1,\quad \text{otherwise}\end{cases},$$

$$\zeta_{\chi}(\omega)=\begin{cases}\zeta_{\chi_{L1}}(\omega),\ p_3'\le\omega\le q_3'\\ \zeta_{\chi_{L2}},q_3'\le\omega\le r_3'\\ \kappa,\quad \omega=r_3'\\ \zeta_{\chi_{U1}}(\omega),\ r_3'\le\omega\le s_3'\\ \zeta_{\chi_{U2}}(\omega),\ s_3'\le\omega\le t_3'\\ 1,\quad \text{otherwise}\end{cases}$$

5.1 Score Function of SVPNN

Assume that $\imath=\langle(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5);(\beta_1,\beta_2,\beta_3,\beta_4,\beta_5);(\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5))\rangle$ is an SVPNN.

Beneficiary degree of truth indicator $=(\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5)/5$.

Hesitation degree of indeterminacy indicator = $(\beta_1+\beta_2+\beta_3+\beta_4+\beta_5)/5$.

Non-beneficiary degree of falsity indicator = $(\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5)/5$.

Chakraborty et al. [159] defined score function $\Gamma(\imath)$ and accuracy function $H(\imath)$ of $\imath$ as follows:

$$\Gamma(\imath)=\frac{1}{3}\left(2+\frac{(\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5)}{5}-\frac{(\beta_1+\beta_2+\beta_3+\beta_4+\beta_5)}{5}\right.$$
$$\left.-\frac{(\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5)}{5}\right)$$

Here, $\Gamma(\imath) \in [0,\ 1]$.

$H(\imath) = \left(\frac{(\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5)}{5} - \frac{(\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5)}{5}\right)$, $H(\imath) \in [-1, 1]$.

5.2 Comparison of SVPNNS

Assume that $\imath_1 = \langle(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5); (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5); (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5))\rangle$ and $\imath_2 = \langle(c_1, c_2, c_3, c_4, c_5); (d_1, d_2, d_3, d_4, d_5); (e_1, e_2, e_3, e_4, e_5))\rangle$ be any two SVPNNs.

Comparison between any two SVPNNs [159] $\imath_1$ and $\imath_2$ is presented as

i. If $\Gamma(\imath_1) > \Gamma(\imath_2)$, then $\imath_1 > \imath_2$;
ii. If $\Gamma(\imath_1) < \Gamma(\imath_2)$, then $\imath_1 < \imath_2$;
iii. If $\Gamma(\imath_1) = \Gamma(\imath_2)$, and

- if $H(\imath_1) > H(\imath_2)$, then $\imath_1 > \imath_2$;
- $H(\imath_1) < H(\imath_2)$, then $\imath_1 < \imath_2$;
- $H(\imath_1) = H(\imath_2)$, then $\imath_1 \approx \imath_2$.

6 Cylindrical Neutrosophic Single-Valued (CNSV) Set

Let W be a space of objects with generic element ω in W. A CNSV set [8] χ in W is presented as: $\chi = \langle \omega, \xi_\chi(\omega),\ \psi_\chi(\omega), \zeta_\chi(\omega) > /,\ \omega \in W \rangle$, where χ $\xi_\chi(\omega)$, $\psi_\chi(\omega), \zeta_\chi(\omega)$ denote the truth MF, the indeterminacy MF, and the falsity MF, respectively.

Here, $(\xi_\chi(\omega))^2 + (\psi_\chi(\omega))^2 \le 1^2,\ \zeta_\chi(\omega) \le 1$.

For convenience, $\chi = (\xi_\chi,\ \psi_\chi, \zeta_\chi)$ is simply defined as a CNSV Number (CNSVN).

6.1 Score Function of CNSVN

For any CNSVN $\chi = (\xi_\chi,\ \psi_\chi, \zeta_\chi)$.

- Beneficiary degree of truth MF = $2(\xi_\chi)^2$
- Indeterminacy degree of indeterminacy MF = $2(\psi)^2$
- Non-beneficiary degree of falsity MF = $2(\zeta_\chi)^2$.

The score function and accuracy function are, respectively, denoted by $\Gamma(\chi)$ and $A(\chi)$ [186] and presented as:

i. $\Gamma(\chi) = \left(2(\xi_\chi)^2 - (\psi_\chi)^2 - (\zeta_\chi)^2\right)$, with $\Gamma(\chi) = [-1, 1]$ and the accuracy function $A(\chi)$ is defined as:

ii. $A(\chi) = \frac{2(\xi_\chi)^2 + (\psi_\chi)^2 + (\zeta_\chi)^2}{2}$, $A(\chi) \in [0, 2]$

6.2 Comparıson of CNSVNs

Assume that $\iota_1 = \left(\xi_{\tau_1}, \psi_{\tau_1}, \zeta_{\tau_1}\right)$ and $\iota_2 = \left(\xi_{\tau_2}, \psi_{\tau_2}, \zeta_{\tau_2}\right)$ are any two CNSVNs. Then, comparison between ι_1 and ι_2 [186] is presented as:

i. If $\Gamma(\iota_1) > \Gamma(\iota_2)$, then $\iota_1 > \iota_2$;
ii. If $\Gamma(\iota_1) < \Gamma(\iota_2)$, then $\iota_1 < \iota_2$;
iii. If $\Gamma(\iota_1) = \Gamma(\iota)$, and

- if $A(\iota_1) > A(\iota_2)$, then $\iota_1 > \iota_2$;
- if $A(\iota_1) < A(\iota_2)$, then $\iota_1 < \iota_2$;
- if $A(\iota_1) = A(\iota_2)$, then $\iota_1 \approx \iota_2$.

7 Neutrosophic Number (NN)

Kandasamy and Smarandache [48, 54] introduced the NN of the structure $\eta = \alpha + \beta\iota$, where α, β denote real or complex numbers, and "ι" denotes the indeterminacy component of η.

An NN η can be presented as $\eta = [\alpha + \beta\iota^l, \alpha + \beta\iota^u]$, $\eta \in N$, N denotes the set having $\forall$ NNs and $\iota \in [\iota^l, \iota^u]$. The interval $\iota \in [\iota^l, \iota^u]$ is called an indeterminate interval.

- When $\beta = 0$, η reduces to crisp number $\eta = \alpha$
- When $\alpha = 0$, then η reduces to the indeterminate number $\eta = \beta I$
- When $\iota^l = \iota^u$, then η reduces to a crisp number.

Assume that $\eta_1 = \alpha_1 + \beta_1\iota$ and $\eta_2 = \alpha_2 + \beta_2\iota$ for $\eta_1, \eta_2 \in$ N and $\iota \in [\iota^l, \iota^u]$ are two NNs. Some basic operational laws [187] for η_1 and η_2 are presented as:

(1) $\iota^2 = \iota$
(2) $\iota.0 = 0$
(3) $\frac{\iota}{\iota} =$ Undefined
(4) $\eta_1 + \eta_2 = \alpha_1 + \alpha_2 + (\beta_1 + \beta_2)\iota = [\alpha_1 + \alpha_2 + (\beta_1 + \beta_2)\iota^l,\ \alpha_1 + \alpha_2 + (\beta_1 + \beta_2)\iota^u]$
(5) $\eta_1 - \eta_2 = \alpha_1 - \alpha_2 + (\beta_1 - \beta_2)\iota = [\alpha_1 - \alpha_2 + (\beta_1 - \beta_2)\iota^l, \alpha_1 - \alpha_2 + (\beta_1 - \beta_2)\iota^u]$
(6) $\eta_1 \times \eta_2 = \alpha_1\alpha_2 - \alpha_2 + (\alpha_1\beta_2 + \alpha_2\beta_1)\iota + \beta_1\beta_2\iota^2 = \alpha_1\alpha_2 - \alpha_2 + (\alpha_1\beta_2 + \alpha_2\beta_1 + \beta_1\beta_2)\iota$

(7) $\frac{\alpha_1+\beta_1\iota}{\alpha_2+\beta_2\iota} = \frac{\alpha_1}{\alpha_2} + \frac{\alpha_2\beta_1-\alpha_1\beta_2}{\alpha_2(\alpha_2+\beta_2)}\iota$; $\alpha_2 \neq 0$, $\alpha_2 \neq -\beta_2$
(8) $\eta_1^2 = \alpha_1^2 + (2\alpha_1\beta_1 + \beta_1^2)\iota$
(9) $\lambda\eta_1 = \lambda\alpha_1 + \lambda\beta_1\iota$

Here, indeterminacy "ι" and the imaginary $i = \sqrt{-1}$ are different concepts. In general, $\iota^p = \iota$ if $p > 0$, and it is undefined if $p < 0$.

7.1 Score Function of NNs

Assume that an *NN is of the form:* $\tau = c + \iota d = [c + d\iota^l, c + d\iota^u]$, *where* c and d *are* not simultaneously zeroes. The score function $\Xi(\chi)$ [54] and the accuracy function $A(\chi)$ [54] are, respectively, presented as

$$\Xi(\tau) = \left|\frac{c + d(\iota^u - \iota^l)}{2\sqrt{c^2 + d^2}}\right|, \Xi(\tau) \in [0, 1]$$

$$A(\tau) = 1 - \exp(-|c + d(\iota^l - \iota^u)|),$$

7.2 Comparison of NNs

Assume that $\eta_1 = c_1 + \iota d_1$ and $\eta_2 = c_2 + \iota d_2$ be any two NNs. Comparison between η_1 and η_2 [54] is presented as follows:

i. If $\Xi(\eta_1) > \Xi(\eta_2)$, then $\eta_1 > \eta_2$;
ii. $\Xi(\eta_1) < \Xi(\eta_2)$, then $\eta_1 < \eta_2$;
iii. If $\Xi(\eta_1) = \Xi(\eta_2)$, and
   - $A(\eta_1) > A(\eta_2)$, then $\eta_1 > \eta_2$;
   - $A(\eta_1) < A(\eta_2)$, then $\eta_1 < \eta_2$;
   - $A(\eta_1) = A(\eta_2)$, then $\eta_1 \approx \eta_2$.

7.3 Neutrosophic Refined Number (NRN)

NRN [188] was defined as:$[p + q_1\iota_1 + q_2\iota_2 + \ldots + q_m\iota_m]$, where $p, q_1, q_2, \ldots, q_m$ denote a real or complex number, and $\iota_1, \iota_2, \ldots, \iota_m$ denote sub-indeterminacies, for $m \geq 1$.

8 Some Applications of Neutrosophic Sets

NSs have been applied in different fields. Few applications are depicted in Table 1.

8.1 Sentiment Analysis

Smarandache et al. [189] studied the similarity measure by defining the words' sentiment scores. In their study, Smarandache et al. [189] dealt with the sentiment characteristics of the words only. They developed a novel word-level similarity measure and found promising results.

8.2 Cosmology

Christianto and Smarandache [190] argued that neutrosophic logic resolved the dispute dealing with the "*beginning and the eternity of the Universe*". In their study, Christianto and Smarandache [190] agreed that "the universe could have both a beginning and an eternal existence" leading to the paradox that "*there might have been a time before time or a beginning of time in time*".

8.3 Neutrosophic Cognitive Map (NCM) for Social Problems

Using the NCM, Devadoss et al. [191] studied to evaluate the impact of playing violent video games among the teenagers (13–18 years) in Chennai. In their study, Devadoss et al. [191] considered nine concepts and presented the outcome of the study by comparing the results derived from Fuzzy Cognitive Map (FCM) and NCM.

8.4 Neutrosophic Strategy to Combat COVID-19

Yasser et al. [192] developed a novel health-fog framework in assisting diagnosis and treatment for COVID-19 patients efficiently based on neutrosophic classifier. The study [192] integrated the information scattered among different medical centres and health organizations to combat with COVID-19.

8.5 Social Network Analysis e-Learning Systems via NSs

Using NSs, Salama et al. [193] integrated social activities in the environment of e-learning and developed a social learning management system. Radwan [194] presented the current trends and challenges in e-learning processes in NS environment.

Table 1 Some applications of NSs and neutrosophic logic

S. No.	Contributors	Contribution of the study
1	Yasser, Twakol, El-Khalek, Samrah, Salama [192]	The study developed a deep learning model to detect COVID-19 patient by employing neutrosophic classifier to extract visual features from volumetric exams.
2	Vasantha, Kandasamy, Smarandache, Devvrat, Ghildiyal [210]	It studied the imaginative play of children by utilizing single valued refined NSs
3	Kandasamy, Vasantha, Obbineni, Smarandache [211]	It analyzed ten political or social datasets of tweets for sentiment analysis using Python and necessary libraries for natural language processing in multi refined NS environment. It presented a more efficient strategy in capturing the opinion of the tweets with best accuracy
4	Vasantha, Kandasamy, Devvrat, Ghildiyal [212]	It studied the imaginative play of children (1–10 years) by employing the NCM
5	Devadoss and Rekha [213]	It analyzed the girls' problems faced by child marriage using the neutrosophic associative FCM
6	Mondal and Pramanik [214]	It analyzed the problems of Hijras in West Bengal using NCMs
7	Pramanik and Chackrabarti [215]	It presented the issues of construction workers in West Bengal using the NCMs
8	Jousselme and Maupin [216]	It described the role of neutrosophy in situation analysis
9	Thiruppathi, Saivaraju, Ravichandran [217]	It studied the suicide problem using combined overlap block NCMs
10	Zafar and Anas [218]	Using NCM, it analyzed the situation of crime in Nigeria
11	William, Devadoss, Sheeba [219]	It analyzed the risk factors of breast cancer
12	Kandasamy and Smarandache [220, 221]	It analyzed the social issues of migrant workers having HIV/AIDS
13	Bernajee [222]	It presented the decision support tool for knowledge based institution
14	Radwan [194]	It described the applications of NS in E-learning
15	Anitha and Gunavathi [195]	The study presented a classification employing musical features
16	Shadrach and Kandasamy [196]	The study provided an early leaf disease diagnosis
17	Pamucar et al. [197]	The study presented a potential energy storage options using fuzzy neutrosophic numbers
18	Ramalingam et al. [223]	The study presented the issues of traffic congestion problem in Indian context

8.6 Raga Classification

Anitha and Gunavathi [195] presented the study of Carnatic raga by classifying all 72 melakarta ragas by utilizing neutrosophic logic and NCM.

8.7 Early Diagnosis of Leaf Ailments

Shadrach and Kandasamy [196] presented a new feature selection strategy in detecting leaf disease using a computer-based method to classify the leaf diseases. In their study, Shadrach and Kandasamy [196] compared the eight existing selection strategies with their developed strategy to demonstrate the capability of their strategy. Their developed strategy [196] obtained 99.8% classification accuracy in selecting 11 characteristics for leaf disease diagnosis.

8.8 Potential Energy Storage Options.

Pamucar et al. [197] developed the neutrosophic Multi-Criteria Decision-Making (MCDM) strategy to evaluate potential energy storage options and conducted a case study in Romania by identifying four criteria and thirteen sub-criteria.

8.9 Neutrosophic Computational Model

Albeanu [198] reviewed the principles of computing and presented some new neutrosophic computational models to identify requirements for software implementation.

8.10 Finance and Economics

Bencze [199] clearly stated that Sukanto Bhattacharya made a significant contribution in neutrosophic research by reflecting the applications of NS-based models dealing with financial economic and social science problems [27, 200–204].

8.11 Conflict Resolution

Bhattacharya et al. [205] presented the arguments in applying the principles of the neutrosophic game theory in order to reflect Israel–Palestine conflict with regard to the goals and strategies of each side. Bhattacharya et al. [205] extended the game theoretic explanation of Plessner [206] and developed the neutrosophic game theoretic model.

Pramanik and Roy [207] presented the arguments in applying the principles of the game theory in order to understand the Indo-Pak conflict over Jammu and Kashmir (J&K) with regard to the goals and strategies of either country. In their study, Pramanik and Roy [207] presented the 2×2 zero-sum game theoretic crisp model of Indo-Pak conflict over J&K by identifying the goals, strategies, and options of either country. Pramanik and Roy [208] also extended the game theoretic crisp model of Pramanik and Roy [207] to neutrosophic game theoretic model to obtain the optimal solution of the ongoing Indo-Pak conflict over J&K.

Deli [209] initiated to study neutrosophic game theory and presented matrix games with simplified neutrosophic payoffs. New research is very important in this area.

8.12 Air Surveillance

Fan et al. [39] proposed the neutrosophic Hough transform (NHT) strategy to deal with the complex surveillance issues. NS is used to characterize the different targets, namely real, false, and uncertain (indeterminate) targets in surveillance environments. They proposed a new neutrosophic Hough transform-based track initiation (NHT-TI) strategy that performs better than modified HT-TI strategy and improved HT-TI strategy.

Air surveillance involves various uncertain factors. In sensor network, uncertain factors may result from unknown target dynamic models, unknown environmental disturbances, the imprecise data processing, and limited performance of sensors [39]. Fan et al. [40] reviewed the NS-based multiple target tracking (MTT) strategies and presented the NS-based MTT strategies and that help in improving the performance.

9 The Extensions of Neutrosophic Sets

NS generalizes the classic set, FS [1] and IFS [7] (see Graphical abstract). NSs have been widely studied, and many extensions have been proposed in the literature. Different applications of NSs are presented in Table 1. The hybrid and extensions of NSs and the contributing authors are shown in Table 2. Currently, there are 82 neutrosophic-related sets derived from neutrosophics in the literature.

10 New Directions

The management of neutrosophic information in human controlled real-world issue appears to be a complex and difficult task. NSs facilitate in handling inconsistency and indeterminacy caused by limited knowledge of the domain experts. This

Table 2 The various hybrid and extensions of NSs

S. No.	Name of the set	Acronym/ Abbreviation	Developed by
1	m-Generalized q-Neutrosophic Set	mGqNS	Saha et al. [224]
2	Generalized Neutrosophic b-Open Set	GNbOS	Das and Pramanik [225]
3	Linguistic neutrosophic set	LNS	Li et al. [141]
4	Plithogenic Set	PS	Smarandache [165]
5	Neutrosophic Crisp Set	NCrS	Salama and Smarandache [226]
6	Interval Neutrosophic Set	INS	Wang et al. [25, 26]
7	Dynamic Interval-Valued Neutrosophic Set	DIVNS	Thong et al. [227]
8	Interval Neutrosophic Linguistic Set	INLS	Ye [142]
9	Single-Valued Neutrosophic Set	SVNS	Wang et al. [9–11]
10	Single Valued Neutrosophic Linguistic Set	SVNLS	Ye [139]
11	Double-Valued Neutrosophic Set	DVNS	Kandasamy [228]
12	Type-2 Single-Valued Neutrosophic Set	T2SVNS	Karaaslan and Hunu [229]
13	Quadripartitioned Single Valued Neutrosophic Set	QSVNS	Chatterjee et al. [164]
14	Triangular Fuzzy Number Neutrosophic Set	TFNNS	Biswas et al. [157]
15	Trapezoidal Fuzzy Neutrosophic Set	TrFNS	Biswas et al. [145]
16	Trapezoidal Neutrosophic Set	TrNS	Ye [146]
17	Single Valued Neutrosophic Trapezoid Linguistic Set	SVNTrLS	Broumi and Smarandache [230]
18	Triangular Neutrosophic Set	TNS	Deli and Subas [231]
19	Simplified Neutrosophic Set	SNS	Ye [41]
20	Simplified Neutrosophic Multiplicative Set	SNMS	Köseoğlu et al. [56]
21	Possibility Simplified Neutrosophic Set	PSNS	Sahin and Liu [232]
22	Neutrosophic Soft Set	NSS	Maji [85]
23	Interval-Valued Neutrosophic Soft Set	IVNSS	Deli [233]
24	Generalized Neutrosophic Soft Set	GNSS	Broumi [234]
25	Generalized Interval Neutrosophic Soft Set	GINSS	Broumi et al. [235]
26	Time-Neutrosophic Soft Set	TNSS	Alkhazaleh [236]

(continued)

Table 2 (continued)

S. No.	Name of the set	Acronym/ Abbreviation	Developed by
27	Time-Neutrosophic Soft Expert Set	TNSES	Ulucay et al. [237]
28	ivnpiv-Neutrosophic Soft Set	ivnpiv-NSS	Deli et al. [92]
29	Interval Valued Neutrosophic Parameterized Soft Set	IVNPSS	Broumi et al. [238]
30	Interval-valued Possibility Quadripartitioned Single Valued Neutrosophic Soft Set	IPQSVNSS	Chatterjee et al. [239]
31	Neutrosophic Valued Linguistic Soft Set	NVLSS	Zhao and Guan [240]
32	Simplified Neutrosophic Uncertain Linguistic Set	SNULS	Tian et al. [143]
33	Bipolar Neutrosophic Set	BNS	Deli et al. [77]
34	Interval Valued Bipolar Neutrosophic Set	IVBNS	Deli et al. [241]
35	Interval Valued Bipolar Fuzzy Weighted Neutrosophic Set	IVBFWNS	Deli et al. [242]
36	Bipolar *Neutrosophic Refined Set*	BNRS	Deli and Subas [161]
37	Complex Neutrosophic Set	CNS	Ali and Smarandache [243]
38	Interval Complex Neutrosophic Set:	ICNS	Ali et al. [244]
39	Complex Neutrosophic Soft Expert Set	CNSES	Al-Quran and Hassan [245]
40	Single-Valued Linguistic Complex Neutrosophic Set	(SVLCNS)	Dat et al. [246]
41	Interval Linguistic Complex Neutrosophic Set	(ILCNS)	Dat et al. [246]
42	Bipolar Complex Neutrosophic Set	BCNS	Broumi et al. [247]
43	Neutrosophic Cubic Set	NCS	Ali et al. [68], Jun et al. [70]
44	Neutrosophic Soft Cubic Set	NSCS	Cruz and Irudayam [248]
45	Possibility Neutrosophic Cubic Set	PNCS	Xue et al. [249]
46	Rough Neutrosophic Set	RNS	Broumi et al. [99, 100]
47	Interval Rough Neutrosophic Set	IRNS	Broumi and Smarandache [101]
48	Single Valued Neutrosophic Rough Set	SVNRS	Yang et al. [102]
49	Single Valued Neutrosophic Multi-Granulation Rough Set	SVNMGRS	Zhang et al.[103]
50	Single Valued Neutrosophic Refined Rough Set	SVNRRS	Bao and Yang [137]

(continued)

Table 2 (continued)

S. No.	Name of the set	Acronym/ Abbreviation	Developed by
51	Interval-Valued Neutrosophic Soft Rough Set	IVNSRS	Broumi and Smarandache [250]
52	Rough Bipolar Neutrosophic Set	RBNS	Pramanik and Mondal [120]
53	Rough Neutrosophic Hyper-Complex Set	RNHCS	Mondal et al. [163]
54	Neutrosophic Hesitant Fuzzy Set	NHFS	Ye [57]
55	Interval Neutrosophic Hesitant Fuzzy Set	INHFS	Ye [251]
56	Triangular Neutrosophic Cubic Linguistic Hesitant Fuzzy	TNCLHF	Fahmi et al. [252]
57	Hesitant Bipolar-Valued Neutrosophic Set	HBVNS	Awang et al. [253]
58	Neutrosophic Refined Set	NRS	Smarandache [126]
59	Single Valued Neutrosophic Multiset	SVNM	Ye and Ye [254]
60	Triple Refined Indeterminate Neutrosophic Set	TRINS	Kandasamy and Smarandache [255]
61	Double Refined Indeterminacy Neutrosophic Set	DRINS	Kandasamy and Smarandache [256]
62	Single-Valued Spherical Neutrosophic Set	SVSp-NS	Smarandache [257]
63	Single-Valued n-Hyper Sphere Neutrosophic Set	SVn-HSNS	Smarandache [257]
64	Neutrosophic Fuzzy Set	NFS	Das et al. [258]
65	Neutrosophic Vague Set	NVS	Alkhazaleh [259]
66	Interval Neutrosophic Vague Set	INVS	Hashim et al. [260]
67	Neutrosophic Vague Soft Set	NVSS	Al-Quran and Hassan [261]
68	Neutrosophic Vague Soft Expert Set	NVSES	Al-Quran and Hassan [262]
69	Vague–valued Possibility Neutrosophic Vague Soft Expert Set	VPNVSPS	Mukherjee [263]
70	Neutrosophic Vague Soft Multiset	NVSM	Al-Quran and Hassan [264]
71	Possibility Neutrosophic Vague Soft Set	PNVSS	Hassan and Al-Quran [265]
72	Neutrosophic Bipolar Vague Set	NBVS	Hussain et al. [266]

(continued)

Table 2 (continued)

S. No.	Name of the set	Acronym/ Abbreviation	Developed by
73	Intuitionistic Neutrosophic Set	InNS	Bhowmik and Pal [267]
74	Intuitionistic Neutrosophic Soft Set	InNSS	Broumi and Smarandache [268]
75	Multi-Valued Neutrosophic Set	MVNS	Wang and Li [60]
76	N-Valued Interval Neutrosophic Set	NVINS	Broumi et al. [269]
77	Probability Multi-Valued Neutrosophic Set	PMVNS	Peng et al. [270]
78	Probability Multi-Valued Linguistic Neutrosophic Set	PMVLNS	Wang and Zhang [271]
79	Normal Neutrosophic Set	NNS	Liu and Teng [272]
80	Multi-Valued Neutrosophic Soft Set	MVNSS	Kamal and Abdullah [273]
81	Multi-Valued Interval Neutrosophic Soft Set	MVINSS	Kamal et al. [274]
82	n-Valued Refined Neutrosophic Soft Set	n-VRNSS	Alkhazaleh [275]

chapter has presented various neutrosophic concepts and tools to deal with neutrosophic information.

To manage neutrosophic information in complex problems, different theoretical strategies [169, 171] have been presented in the current neutrosophic literature. Some weaknesses of the neutrosophic studies are highlighted.

- Since the introduction of SVNS [11], many extensions and versions of NSs have been proposed. But some of the extensions of NSs are debatable with respect to their usefulness. These extensions should be able to solve real problems with uncertainty, inconsistency, and indeterminacy. Theoretical or practical dimensions of the extensions must be justified.
- Some papers with the same title and almost same content [2–6, 9–11, 99, 100] have been published in different journals leading to the confusion, self-plagiarism and the waste of time to find differences between them.
- Too many NSs and decision-making strategies based on these sets have been proposed in the literature without presenting a commanding justification of their applications in real-life problems. It appears that a commanding justification of their necessity and applicability are necessary.
- Researchers utilize different notations for presenting concepts, tools, and extensions to deal with neutrosophic information.
- The increase in volume, uncertainty, inconsistency, falsity, indeterminacy, and incompleteness in data reflects currently a major challenge. The current scenario demands the scientific analysis and new development of neutrosophic

frameworks that are capable of modelling, clustering, and data fusion problems in different fields of scientific study. Uncertainty, falsity and indeterminacy in prices are the key aspects in economic activities. So, new neutrosophic research should address the diverse topics such as stock trending analysis [276], performance of stock market [277], finance, economics and politics [27, 199–204], conflict resolution [208], air surveillance, and multiple target tracking strategy [37, 39].

- Neutrosophic game theoretic model [209] must be further studied.
- A new trend in research appears as the utilization of the neutrosophic theoretical models to realistic problems. Neutrosophic models should be a new and provide solutions to the problems, which cannot be solved by developed strategies in the literature.
- Since vague sets [278] are IFSs, and IFS is equivalent interval FS [279], the relations between neutrosophic FS [258], vague NS [259], and intuitionistic NS [267] should be deeply investigated.
- In upcoming 30 years, NS and its hybrid sets will highly be useful in artificial intelligence, automation, cybernetics, data analysis, engineering management science, mobile ad hoc network, neurosciences, operations research, interdisciplinary applications, multidisciplinary science, weather forecasting, etc.

11 Conclusion

Uncertainty, indeterminacy and inconsistency usually get involved in many human controlled complex real-world problems. NSs and their various extensions offer successful results in dealing with different neutrosophic decision-making problems. Much attentions have been given to some of them that manage neutrosophic situations, which often appear when indeterminacy must be dealt with. These new strategies have attracted the great attention of the investigators who deeply studied the diverse neutrosophic concepts, diverse hybrid extensions, similarity measures, and aggregation operators to deal with neutrosophic information.

The chapter has presented some directions and considerations of future researches that should be considered in the coming NS-based proposals. It has also recognizez that there exist many avenues of research in NSs and their extensions. It is to be noted that the results of this chapter offer a current overview of NSs and SVNSs. However, NSs will evolve and possibly develop in the future according to new ideas and topics that will dominate the current neutrosophic research arena.

Core Messages

- NS offers a natural foundation for dealing with the neutrosophic phenomena mathematically that exist widely in the real world.
- For realistic decision-making problems, the neutrosophic set is a very promising mathematical tool that can dominate the other mathematical tools for making pragmatic and rational decision-makıng in a complex and uncertain environment.
- NSs are capable of handling uncertainty, inconsistency, indeterminacy, and incompleteness and that is why they attract all branches of knowledge.
- NS is a promising mathematical tool for integrated science.

References

1. Zadeh LA (1965) Fuzzy sets. Inf Control 5:338–353
2. Smarandache F (1998) Neutrosophy: neutrosophic probability, set, and logic: analytic synthesis & synthetic analysis. American Research Press, Rehoboth
3. Smarandache F (1999) Linguistic paradoxes and tautologies. Libertas Mathematica, University of Texas at Arlington XIX:143–154
4. Smarandache F (2005) Neutrosophic set–a generalization of the intuitionistic fuzzy set. Int J Pure Appl Math 24(3):287–297
5. Smarandache F (2006) Neutrosophic set-a generalization of the intuitionistic fuzzy set. In: 2006 IEEE international conference on granular computing. IEEE, New York, pp 38–42
6. Smarandache F (2010) Neutrosophic set–a generalization of the intuitionistic fuzzy set. J Defense Resources Manage 1(1):107–116
7. Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20(1):87–96
8. Smarandache F (2019) Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, spherical fuzzy set, and q-rung orthopair fuzzy set, while neutrosophication is a generalization of regret theory, grey system theory, and three-ways decision (revisited). J New Theory 29:1–31
9. Wang H, Smarandache F, Zhang Y, Sunderraman R (2005) Single valued neutrosophic sets. In: Proceedings of 10th 476 International conference on fuzzy theory and technology
10. Wang H, Smarandache F, Zhang Y, Sunderraman R (2010) Single valued neutrosophic sets. Rev Air Force Acad 1:10–14
11. Wang H, Smarandache F, Zhang Y, Sunderraman R (2010) Single valued neutrosophic sets. Multispace Multistructure 4:410–413
12. Sodenkamp M (2013) Models, methods and applications of group multiple-criteria decision analysis in complex and uncertain systems. Dissertation, University of Paderborn, Germany
13. Ye J (2013) Multicriteria decision-making method using the correlation coefficient under single-value neutrosophic environment. Int J Gen Syst 42(4):386–394
14. Liu P, Wang Y (2014) Multiple attribute decision-making method based on single-valued neutrosophic normalized weighted Bonferroni mean. Neural Comput Appl 25(7–8):2001–2010
15. Kharal A (2014) A neutrosophic multi-criteria decision making method. New Mathe Natural Comput 10(2):143–162

16. Biswas P, Pramanik S, Giri BC (2014) Entropy based grey relational analysis method for multi-attribute decision making under single valued neutrosophic assessments. Neutrosophic Sets Syst 2:102–110
17. Biswas P, Pramanik S, Giri BC (2014) A new methodology for neutrosophic multi-attribute decision-making with unknown weight information. Neutrosophic Sets Systs 3:42–50
18. Mondal K, Pramanik S (2015) Neutrosophic decision making model of school choice. Neutrosophic Sets Syst 7:62–68
19. Mondal K, Pramanik S (2015) Neutrosophic tangent similarity measure and its application to multiple attribute decision making. Neutrosophic Sets Syst 9:80–87
20. Şahin R, Liu P (2015) Maximizing deviation method for neutrosophic multiple attribute decision making with incomplete weight information. Neural Comput Appl 27(7):2017–2029
21. Pramanik S, Biswas P, Giri BC (2017) Hybrid vector similarity measures and their applications to multi-attribute decision making under neutrosophic environment. Neural Comput Appl 28(5):1163–1176
22. Biswas P, Pramanik S, Giri BC (2016) TOPSIS method for multi-attribute group decision making under single-valued neutrosophic environment. Neural Comput Appl 27(3):727–737
23. Biswas P, Pramanik S, Giri BC (2019) Non-linear programming approach for single-valued neutrosophic TOPSIS method. New Math Natural Comput 15(2):307–326
24. Biswas P, Pramanik S, Giri, BC (2019) Neutrosophic TOPSIS with group decision making. In: Fuzzy multi-criteria decision-making using neutrosophic sets. Springer, berlin, pp 543–585
25. Wang H, Madiraju P, Zhang YQ, Sunderraman R (2005) Interval neutrosophic sets. Int J Appl Math Stat 3:1–18
26. Wang H, Smarandache F, Zhang YQ, Sunderraman R (2005) Interval neutrosophic sets and logic: theory and applications in computing. Arizona, Hexis
27. Bhattacharya S (2005) Neutrosophic information fusion applied to the options market. Investment Manage Fin Innov 1:139–145
28. Khoshnevisan M, Bhattacharya S (2003) Neutrosophic information fusion applied to financial market. In: Sixth international conference of information fusion, 2003. Proceedings of the, vol 2. IEEE, New York, pp 1252–1257
29. Guo Y, Cheng HD, Zhang Y, Zhao W (2008) A new neutrosophic approach to image denoising. In: Proceedings of the 11th joint conference on information sciences. Atlantis Press, pp 1–6
30. Guo Y, Cheng HD (2009) New neutrosophic approach to image segmentation. Pattern Recogn 42(5):587–595
31. Ju W (2011) Novel application of neutrosophic logic in classifiers evaluated under region-based image categorization system. All Graduate Theses and Dissertations. 887. https://digitalcommons.usu.edu/etd/887
32. Koundal D, Gupta S, Singh S (2016) Applications of neutrosophic sets in medical image denoising and segmentation. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and application. Pons Editions, Brussels, pp 257–275
33. Ye J (2014) Single valued neutrosophic minimum spanning tree and its clustering method. J Intell Syst 23(3):311–324
34. Dutta AK, Biswas R, Al-Arifi NS (2011) A study of neutrosophic technology to retrieve queries in relational database. Int J Comput Sci Emerg Technol 2(1):133–138
35. Arora M, Pandey US (2011) Supporting queries with imprecise constraints in total neutrosophic data bases. Global J Sci Front Res 11(8):67–72
36. Patro SK (2016) On a model of love dynamics: a neutrosophic analysis. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and application. Pons Editions, Brussels, pp 279–287

37. Hu K, Ye J, Fan E, Shen S, Huang L, Pi J (2017) A novel object tracking algorithm by fusing color and depth information based on single valued neutrosophic cross-entropy. J Intell Fuzzy Syst 32:1775–1786
38. Ye J (2017) Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine. Soft Comput 21(3):817–825
39. Fan E, Hu K, Li X (2019) Review of neutrosophic-set-theory-based multiple-target tracking methods in uncertain situations. In: 2019 IEEE international conference on artificial intelligence and computer applications (ICAICA). IEEE, New York, pp 19–27
40. Fan E, Xie W, Pei J, Hu K, Li X (2018) Neutrosophic Hough transform-based track initiation method for multiple target tracking. IEEE Access 6:16068–16080
41. Ye J (2014) Vector similarity measures of simplified neutrosophic sets and their application in multicriteria decision making. Int J Fuzzy Syst 16(2):204–211
42. Peng JJ, Wang JQ, Zhang HY, Chen XQ (2014) An outranking approach for multi-criteria decision-making problems with simplified neutrosophic sets. Appl Soft Comput 25:336–346
43. Peng JJ, Wang JQ, Wang ZHY, Chen XH (2016) Simplified neutrosophic sets and their applications in multi-criteria group decision-making problems. Int J Syst Sci 47(10):2342–2358
44. Wu XH, Wang JQ, Peng JJ, Chen XH (2016) Cross-entropy and prioritized aggregation operator with simplified neutrosophic sets and their application in multi-criteria decision-making problems. Int J Fuzzy Syst 18:1104–1116
45. Şahin R, Küçük GD (2018) Group decision making with simplified neutrosophic ordered weighted distance operator. Math Methods Appl Sci 41(12):4795–4809
46. Du S, Yong R, Ye J (2020) Subtraction operational aggregation operators of simplified neutrosophic numbers and their multi-attribute decision making approach. Neutrosophic Sets Syst 33:157–167
47. Ye J (2015) Improved cosine similarity measures of simplified neutrosophic sets for medical diagnoses. Artif Intell Med 63(3):171–179
48. Kandasamy WBV, Smarandache F (2003) Fuzzy cognitive maps and neutrosophic cognitive maps. ProQuest Information and Learning, Ann Arbor
49. Pramanik S, Roy R, Roy TK (2017) Teacher selection strategy based on bidirectional projection measure in neutrosophic number environment. In: Smarandache F, Basset MA, Chang V (eds) Neutrosophic operational research, vol II. Pons Edition, Brussels, pp 29–53
50. Ye J (2016) Multiple-attribute group decision-making method under a neutrosophic number environment. J Intell Syst 25(3):377–386
51. Ye J (20117). Bidirectional projection method for multiple attribute group decision making with neutrosophic number. Neural Comput Appl 28:1021–1029
52. Liu P, Liu X (2018) The neutrosophic number generalized weighted power averaging operator and its application in multiple attribute group decision making. Int J Mach Learn Cybern 9:347–358
53. Zheng E, Teng F, Liu P (2017) Multiple attribute group decision-making method based on neutrosophic number generalized hybrid weighted averaging operator. Neural Comput Appl 28:2063–2074
54. Mondal K, Pramanik S, Giri BC, Smarandache F (2018) NN-Harmonic mean aggregation operators-based MCGDM strategy in a neutrosophic number environment. Axioms 7(1):12
55. Du S, Ye J, Yong R, Zhang F (2020) Simplified neutrosophic indeterminate decision making method with decision makers' indeterminate ranges. J Civ Eng Manag 26(6):590–598
56. Köseoğlu A, Şahin R, Merdan M (2019) A simplified neutrosophic multiplicative set-based TODIM using water-filling algorithm for the determination of weights. Expert Syst. https://doi.org/10.1111/exsy.12515
57. Ye J (2015) Multiple-attribute decision-making method under a single-valued neutrosophic hesitant fuzzy environment. J Intell Syst 24(1):23–36
58. Torra V (2010) Hesitant fuzzy sets. Int J Intell Syst 25(6):529–539

59. Zhu B, Xu Z, Xia M (2012) Dual hesitant fuzzy sets. J Appl Math. Article ID 879629. https://doi.org/10.1155/2012/879629
60. Wang JJ, Li XE (2015) TODIM method with multi-valued neutrosophic sets. Control Decis 30(6):1139–1142
61. Sahin R, Liu P (2016) Distance and similarity measure for multiple attribute with single – valued neutrosophic hesitant fuzzy information. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, Brussels, pp 35–54
62. Sahin R, Liu P (2017) Correlation coefficient of single-valued neutrosophic hesitant fuzzy sets and its applications in decision making. Neural Comput Appl 28(6):1387–1395
63. Biswas P, Pramanik S, Giri BC (2016) GRA method of multiple attribute decision making with single valued neutrosophic hesitant fuzzy set information. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, Brussels, pp 55–63
64. Biswas P, Pramanik S, Giri BC (2016) Some distance measures of single valued neutrosophic hesitant fuzzy sets and their applications to multiple attribute decision making. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, Brussels, pp 55–63
65. Biswas P, Pramanik S, Giri, BC (2019) NH-MADM strategy in neutrosophic hesitant fuzzy set environment based on extended GRA. Informatica 30(2):1–30
66. Peng JJ, Wang J (2015) Multi-valued neutrosophic sets and its application in multi-criteria decision-making problems. Neutrosophic Sets Syst 10:3–17
67. Peng JJ, Wang J, Wu X, Wang J, Chen X (2014) Multi-valued neutrosophic sets and power aggregation operators with their applications in multi-criteria group decision-making problems. Int J Comput Intell Syst 8(2):345–363
68. Ali M, Deli I, Smarandache F (2016) The theory of neutrosophic cubic sets and their applications in pattern recognition. J Intell Fuzzy Syst 30(4):1957–1963
69. Jun YB, Kim CS, Yang KO (2012) Cubic sets. Ann Fuzzy Math Inform 4(3):83–98
70. Jun YB, Smarandache F, Kim CS (2017) Neutrosophic cubic sets. New Math Natural Comput 13(1):41–54
71. Banerjee D, Giri BC, Pramanik S, Smarandache F (2017) GRA for multi attribute decision making in neutrosophic cubic set environment. Neutrosophic Sets Syst 15:60–69
72. Pramanik S, Dey PP, Giri BC, Smarandache F (2017) An extended TOPSIS for multi-attribute decision making problems with neutrosophic cubic information. Neutrosophic Sets Syst 17:70–79
73. Lu Z, Ye J (2017) Cosine measures of neutrosophic cubic sets for multiple attribute decision-making. Symmetry 9(7):121
74. Zhan J, Khan M, Gulistan M, Ali A (2017) Applications of neutrosophic cubic sets in multi-criteria decision-making. Int J Uncertain Quantif 7(5):377–394
75. Pramanik S, Dalapati S, Alam S, Roy TK (2017) NC-TODIM-based MAGDM under a neutrosophic cubic set environment. Information 8(4):149
76. Pramanik S, Dalapati S, Alam S, Roy TK, Smarandache F (2017) Neutrosophic cubic MCGDM method based on similarity measure. Neutrosophic Sets Syst 16:44–56
77. Deli I, Ali M, Smarandache F (2015) Bipolar neutrosophic sets and their applications based on multicriteria decision making problems. In: International conference on advanced mechatronic systems (ICAMechs), pp 249–254
78. Lee KM (2000) Bipolar-valued fuzzy sets and their operations. In: Proceedings of the international conference on intelligent technologies, Bangkok, Thailand, pp 307–312
79. Uluçay V, Del I, Şahin M (2018) Similarity measures of bipolar neutrosophic sets and their application to multiple criteria decision making. Neural Comput Appl 29:739–748
80. Dey PP, Pramanik S, Giri BC (2016) TOPSIS for solving multi-attribute decision making problems under bi-polar neutrosophic environment. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, Brussels, pp 65–77

81. Pramanik S, Dey PP, Giri BC, Smarandache F (2017) Bipolar neutrosophic projection based models for solving multi-attribute decision making problems. Neutrosophic Sets Syst 15:70–79
82. Wang L, Zhang HY, Wang JQ (2018) Frank Choquet Bonferroni mean operators of bipolar neutrosophic sets and their application to multi-criteria decision-making problems. Int J Fuzzy Syst 20(1):13–28
83. Fan C, Ye J, Feng S, Fan E, Hu K (2019) Multi-criteria decision-making method using Heronian mean operators under a bipolar neutrosophic environment. Mathematics 7(1):97
84. Jamil M, Abdullah S, Khan MY, Smarandache F, Ghani F (2019) Application of the bipolar neutrosophic Hamacher averaging aggregation operators to group decision making: an illustrative example. Symmetry 11(5):698
85. Maji PK (2013) Neutrosophic soft set. Ann Fuzzy Math Inform 5(1):157–168
86. Molodstov D (1999) Soft set theory-first results. Comput Math Appl 37:19–31
87. Deli I, Broumi S (2015) Neutrosophic soft matrices and NSM-decision making. J Intell Fuzzy Syst 28(5):2233–2241
88. Dey PP, Pramanik S, Giri BC (2016) Neutrosophic soft multi-attribute decision making based on grey relational projection method. Neutrosophic Sets Syst 11:98–106
89. Şahin R, Küçük A (2014) On similarity and entropy of neutrosophic soft sets. J Intell Fuzzy Syst 27(5):2417–2430
90. Deli I, Broumi S, Ali M (2014) Neutrosophic soft multi-set theory and its decision making. Neutrosophic Sets Syst 5:65–76
91. Jana C, Pal M (2019) A robust single-valued neutrosophic soft aggregation operators in multi-criteria decision making. Symmetry 11(1):110. https://doi.org/10.3390/sym11010110
92. Deli I, Erasla S, Çağman N (2018) ivnpiv-Neutrosophic soft sets and their decision making based on similarity measure. Neural Comput Appl 29:187–203
93. Dey PP, Pramanik S, Giri BC (2015) Generalized neutrosophic soft multi-attribute group decision making based on TOPSIS. Crit Rev 11:41–55
94. Dey PP, Pramanik S, Giri BC (2016) Neutrosophic soft multi-attribute group decision making based on grey relational analysis method. J New Res Sci 10:25–37
95. Pramanik S, Dalapati S (2016) GRA based multi criteria decision making in generalized neutrosophic soft set environment. Global J Eng Sci Res Manage 3(5):153–169
96. Das S, Kumar S, Kar S, Pal T (2017) Group decision making using neutrosophic soft matrix: an algorithmic approach. J King Saud Univ – Comput Inf Sci 31(4):459–546
97. Ali M, Son LH, Deli I, Tien ND (2017) Bipolar neutrosophic soft sets and applications in decision making. J Intell Fuzzy Syst 33(6):4077–4087
98. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:341–356
99. Broumi S, Smarandache F, Dhar M (2014) Rough neutrosophic sets. Italian J Pure Appl Math 32:493–502
100. Broumi S, Smarandache F, Dhar M (2014) Rough neutrosophic sets. Neutrosophic Sets Syst 3:60–66
101. Broumi S, Smarandache F (2015) Interval neutrosophic rough set. Neutrosophic Sets Syst 7:23–31
102. Yang HL, Zhang CL, Guo ZL, Liu YL, Liao X (2017) A hybrid model of single valued neutrosophic sets and rough sets: single valued neutrosophic rough set model. Soft Comput 21:6253–6267
103. Zhang C, Zhai Y, Li D, Mu Y (2016) Steam turbine fault diagnosis based on single-valued neutrosophic multigranulation rough sets over two universes. J Intell Fuzzy Syst 31(6):2829–2837
104. Qian YH, Liang JY (2006) Rough set method based on multi-granulations. In: 2006 5th IEEE international conference on cognitive informatics, vol 1. IEEE, New York, pp 297–304
105. Jiao L, Yang HL, Li SG (2020) Three-way decision based on decision-theoretic rough sets with single-valued neutrosophic information. Int J Mach Learn Cybern 11:657–665

106. Zhao XR, Hu BQ (2020) Three-way decisions with decision theoretic rough sets in multiset-valued information tables. Inf Sci 507:684–699
107. Mondal K, Pramanik S (2015) Rough neutrosophic multi-attribute decision-making based on grey relational analysis. Neutrosophic Sets Syst 7:8–17
108. Mondal K, Pramanik S (2015) Rough neutrosophic multi-attribute decision-making based on rough accuracy score function. Neutrosophic Sets Syst 8:14–21
109. Mondal K, Pramanik S, Giri BC (2019) Rough neutrosophic aggregation operators for multi-criteria decision-making. Fuzzy multi-criteria decision-making using neutrosophic set. Springer, Cham, pp 79–105
110. Mondal K, Pramanik S, Smarandache F (2016) Several trigonometric Hamming similarity measures of rough neutrosophic sets and their applications in decision making. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and application. Pons Editions, Brussels, pp 93–103
111. Mondal K, Pramanik S, Smarandache F (2016) Multi-attribute decision making based on rough neutrosophic variational coefficient similarity measure. Neutrosophic Sets Syst 13:3–17
112. Mondal K, Pramanik S (2015) Decision making based on some similarity measures under interval rough neutrosophic environment. Neutrosophic Sets Syst 10:46–57
113. Pramanik S, Mondal K (2015) Cosine similarity measure of rough neutrosophic sets and its application in medical diagnosis. Global J Adv Res 2(1):212–220
114. Pramanik S, Mondal K (2015) Some rough neutrosophic similarity measure and their application to multi attribute decision making. Global J Eng Sci Res Manage 2(7):61–74
115. Pramanik S, Mondal K (2015) Cotangent similarity measure of rough neutrosophic sets and its application to medical diagnosis. J New Theory 4:90–102
116. Mondal K, Pramanik S, Smarandache F (2016) Rough neutrosophic TOPSIS for multi-attribute group decision making. Neutrosophic Sets Syst 13:105–117
117. Pramanik S, Roy R, Roy TK, Smarandache F (2017) Multi criteria decision making using correlation coefficient under rough neutrosophic environment. Neutrosophic Sets Syst 17:29–36
118. Pramanik S, Roy R, Roy TK (2018) Multi criteria decision making based on projection and bidirectional projection measures of rough neutrosophic sets. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, Brussels, pp 175–187
119. Roy R, Pramanik S, Roy TK (2020) Interval rough neutrosophic TOPSIS strategy for multi-attribute decision making. In: Abdel-Basset M, Smarandache F (eds) Neutrosophic sets in decision analysis and operations research. IGI Global, pp 98–118
120. Pramanik S, Mondal K (2016) Rough bipolar neutrosophic set. Global J Eng Sci Res Manage 3(6):71–81
121. Pramanik S (2020) Rough neutrosophic set: an overview. In:Smarandache F, Broumi S (eds) Neutrosophic theories in communication, management and information technology. Nova Science Publishers, pp 275–311
122. Yager RR (1986) On the theory of bags. Int J General Syst 13(1):23–37
123. Knuth D (1981) The art of computer programming, vol 2: semi numerical Algorithms. Addison-Wesley, Reading, MA
124. Manna Z, Waldinger R (1985) The logical bases for computer programming, vol 1: deductive reasoning. Addison-Wesley, Reading, MA
125. Shinoj TK, John SJ (2012) Intuitionistic fuzzy multisets and its application in medical diagnosis. World Acad Sci Eng Technol 6(1):1418–1421
126. Smarandach F (2013) n-Valued refined neutrosophic logic and its applications in Physics. Prog Phys 4:143–146
127. Broumi S, Smarandache F (2014) Neutrosophic refined similarity measure based on cosine function. Neutrosophic Sets Syst 6:42–48

128. Broumi S, Deli I (2014) Correlation measure for neutrosophic refined sets and its application in medical diagnosis. Palestine J Math 3:11–19
129. Deli I, Broumi S, Smarandache F (2015) On neutrosophic refined sets and their applications in medical diagnosis. J New Theory 6:88–89
130. Mondal K, Pramanik S (2015) Neutrosophic refined similarity measure based on tangent function and its application to multi attribute decision making. J New Theory 8:41–50
131. Mondal K, Pramanik S (2015) Neutrosophic refined similarity measure based on cotangent function and its application to multi-attribute decision making. Global J Adv Res 2(2):486–494
132. Pramanik S, Banerjee D, Giri BC (2016) TOPSIS approach for multi attribute group decision making in refined neutrosophic environment. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications. Pons Editions, 79–91
133. Pramanik S, Banerjee D, Giri BC (2016) Multi-criteria group decision making model in neutrosophic refined set and its application. Global J Eng Sci Res Manage 3(6):12–18
134. Ye J, Smarandache F (2016) Similarity measure of refined single-valued neutrosophic sets and its multicriteria decision making method. Neutrosophic Sets Syst 12:41–44
135. Chen J, Ye J, Du S (2107) Vector similarity measures between refined simplified neutrosophic sets and their multiple attribute decision-making method. Symmetry 9(8):153
136. Pramanik S, Dey PP, Giri BC (2018) Hybrid vector similarity measure of single valued refined neutrosophic sets to multi-attribute decision making problems. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications, vol 2. Pons Editions, pp 156–174
137. Bao YL, Yang HL (2017) On single valued neutrosophic refined rough set model and its application. J Intell Fuzzy Syst 33:1235–1248
138. Bao YL, Yang HL (2019) On single valued neutrosophic refined rough set model and its application. In: Kahraman C, Otay I (eds) fuzzy multi-criteria decision-making using neutrosophic sets. Studies in fuzziness and soft computing, vol 369. Springer, Berlin, pp 107–143
139. Ye J (2015) An extended TOPSIS method for multiple attribute group decision making based on single valued neutrosophic linguistic numbers. J Intell Fuzzy Syst 28:247–255
140. Wang JQ, Li HB (2010) Multi-criteria decision making based on aggregation operators for intuitionistic linguistic fuzzy numbers. Control Decis 25(10):1571–1574
141. Li YY, Zhang H, Wang JQ (2017) Linguistic neutrosophic sets and their application in multicriteria decision-making problems. Int J Uncertain Quantif 7(2):135–154
142. Ye J (2014) Some aggregation operators of interval neutrosophic linguistic numbers for multiple attribute decision making. J Intell Fuzzy Syst 27:2231–2241
143. Tian ZP, Wang J, Zhang HY, Chen XH, Wang JQ (2016) Simplified neutrosophic linguistic normalized weighted Bonferroni mean operator and its application to multi-criteria decision-making problems. Filomat 30(12):3339–3360
144. Xu Z (2005) Deviation measures of linguistic preference relations in group decision making. Omega 33(3):249–254
145. Biswas P, Pramanik S, Giri BC (2014) Cosine similarity measure based multi-attribute decision-making with trapezoidal fuzzy neutrosophic numbers. Neutrosophic Sets Syst 8:46–56
146. Ye J (2015) Trapezoidal neutrosophic set and its application to multiple attribute decision-making. Neural Comput Appl 26:1157–1166
147. Deli I, Subas Y (2017) A ranking method of single valued neutrosophic numbers and its application to multi-attribute decision making problems. Int J Mach Learn Cybern 8 (4):1309–1322
148. Biswas P, Pramanik S, Giri BC (2018) Multi-attribute group decision making based on expected value of neutrosophic trapezoidal numbers. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and applications, vol 2. Pons Editions, pp 103–124

149. Biswas P, Pramanik S, Giri BC (2018) TOPSIS strategy for multi-attribute decision making with trapezoidal neutrosophic numbers. Neutrosophic Sets Syst 19:29–39
150. Pramanik S, Mallick R (2018) VIKOR based MAGDM strategy with trapezoidal neutrosophic numbers. Neutrosophic Sets Syst 22:118–130
151. Pramanik S, Mallick R (2019) TODIM strategy for multi-attribute group decision making in trapezoidal neutrosophic number environment. Complex Intell Syst 5(4):379–389
152. Pramanik S, Mallick R (2020) Extended GRA-based MADM strategy with single-valued trapezoidal neutrosophic numbers. In: Abdel-Basset M, Smarandache F (eds) Neutrosophic sets in decision analysis and operations research. IGI Global, pp 150–179
153. Jana C, Pal M, Karaaslan F, Wang JQ (2020) Trapezoidal neutrosophic aggregation operators and their application to the multi-attribute decision-making process. Scientia Iranica. Transaction E, Ind Eng 27(3):1655–1673
154. Biswas P, Pramanik S, Giri BC (2016) Value and ambiguity index based ranking method of single-valued trapezoidal neutrosophic numbers and its application to multi-attribute decision making. Neutrosophic Sets Syst 12:127–138
155. Liang RX, Wang JQ, Zhang HY (2018) A multi-criteria decision making method based on single valued trapezoidal neutrosophic preference relation with complete weight information. Neural Comput Appl 30:3383–3398
156. Biswas P, Pramanik S, Giri BC (2018) Distance measure based MADM strategy with interval trapezoidal neutrosophic numbers. Neutrosophic Sets Syst 19:40–46
157. Biswas P, Pramanik GBC (2016) Aggregation of triangular fuzzy neutrosophic set information and its application to multi-attribute decision making. Neutrosophic Sets Syst 12:20–40
158. Abdel-Basset M, Saleh M, Gamal A, Smarandache F (2019) An approach of TOPSIS technique for developing supplier selection with group decision making under type-2 neutrosophic number. Appl Soft Comput 77:438–452
159. Chakraborty A, Broumi S, Singh PK (2019) Some properties of pentagonal neutrosophic numbers and its applications in transportation problem environment. Neutrosophic Sets Syst 28:200–215
160. Karaaslan F (2018) Gaussian single-valued neutrosophic numbers and its application in multi-attribute decision making. Neutrosophic Sets Syst 22:101–117
161. Deli I, Subas Y (2016) Bipolar neutrosophic refined sets and their applications in medical diagnosis. In: International conference on natural science and engineering (ICNASE'16), pp 1121–1132
162. Mondal K, Pramanik S (2015) Tri-complex rough neutrosophic similarity measure and its application in multi-attribute decision making. Crit Rev 11:26–40
163. Mondal K, Pramanik S, Smarandache F (2016) Rough neutrosophic hyper-complex set and its application to multi-attribute decision making. Crit Rev 13:111–126
164. Chatterjee R, Majumdar P, Samanta SK (2016) On some similarity measures and entropy on quadripartitioned single valued neutrosophic sets. J Intell Fuzzy Syst 30(4):2475–2485
165. Smarandache F (2017) Plithogeny, plithogenic set, logic, probability, and statistics. Pons asbl, Brussels
166. Smarandache F (2018) Plithogenic set: an extension of crisp, fuzzy, intuitionistic fuzzy, and neutrosophic sets–revisited. Neutrosophic Sets Syst 21:153–166
167. Smarandache F (2016) Operators on single valued neutrosophic oversets, neutrosophic undersets, and neutrosophic offsets. J Math Inf 5:63–67
168. Smarandache F (2016) Neutrosophic overset, neutrosophic underset, neutrosophic offset, similarly for neutrosophic over-/under-/off logic, probability, and statistic. Pons Editions
169. Smarandache F, Pramanik S (eds) (2016) New trends in neutrosophic theory and applications. Pons Editions
170. El-Hefenawy N, Metwally MA, Ahmed ZM, El-Henawy IM (2016) A review on the applications of neutrosophic sets. J Comput Theor Nanosci 13(1):936–944

171. Smarandache F, Pramanik S (eds) (2018) New trends in neutrosophic theory and applications, vol 2. Pons Editions
172. Khan M, Son LH, Ali M, Chau HT, Na NTN, Smarandache F (2018) Systematic review of decision making algorithms in extended neutrosophic sets. Symmetry 10(8):314
173. Pramanik S, Mallick R, Dasgupta A (2018) Contributions of selected Indian researchers to multi-attribute decision making in neutrosophic environment. Neutrosophic Sets Syst 20:108–131
174. Nguyen GN, Son LH, Ashour AS, Dey N (2019) A survey of the state-of-the-arts on neutrosophic sets in biomedical diagnoses. Int J Mach Learn Cybern 10:1–13. https://doi.org/10.1007/s13042-017-0691-7
175. Akram M, Shum, KP (2020) A survey on single-valued neutrosophic K−algebras. J Math Res Appl 40(3). https://doi.org/10.3770/j.issn:2095-2651.2020.03.000
176. Peng X, Dai J (2020) A bibliometric analysis of neutrosophic set: two decades review from 1998 to 2017. Artif Intell Rev 53(1):199–255
177. Muzaffar A, Nafis MT, Sohail SS (2020) Neutrosophy logic and its classification: an overview. Neutrosophic Sets Syst 35:239–251
178. Sahin R (2014) Multi-criteria neutrosophic decision making method based on score and accuracy functions under neutrosophic environment. http://arxiv.org/abs/1412.5202. Accessed on 7 July 2020
179. Nancy GH (2016) An improved score function for ranking neutrosophic sets and its application to decision-making process. Int J Uncertain Quantif 6(5):377–385
180. Chi P, Liu P (2013) An extended TOPSIS method for the multiple attribute decision making problems based on interval neutrosophic set. Neutrosophic Sets Syst 1:63–70
181. Kutlu Gündoğdu F, Kahraman C (2019) Spherical fuzzy sets and spherical fuzzy TOPSIS method. J Intell Fuzzy Syst 36(1):337–352
182. Smarandache F (2017) Neutrosophic perspectives: triplets, duplets, multisets, hybrid operators, modal logic, hedge algebras, and applications, 2nd edn. Pons Publishing House
183. Deli I, Subas Y (2014) Single valued neutrosophic numbers and their applications to multicriteria decision making problem. viXra preprint viXra 14120012
184. Wang JH, Hao JY (2006) A new version of 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans Fuzzy Syst 4(3):435–445
185. Wang YM (2009) Centroid defuzzification and the maximizing set and minimizing set ranking based on alpha level sets. Comput Ind Eng 57(1):228–236
186. Chakraborty A, Mondal SP, Alam S, Mahata A (2020) Cylindrical neutrosophic single-valued number and its application in networking problem, multi-criterion group decision-making problem and graph theory. CAAI Trans Intell Technol 5(2):68–77
187. Smarandache F (2014) Introduction to neutrosophic statistics. Sitech & Education Publishing, Columbus
188. Smarandache F (2105) Neutrosophic quadruple numbers, refined neutrosophic quadruple numbers, absorbance law, and the multiplication of neutrosophic quadruple numbers. Neutrosophic Sets Syst 10:96–98
189. Smarandache F, Colhon M, Vlăduţescu Ş, Negrea X (2019) Word-level neutrosophic sentiment similarity. Appl Soft Comput 80:167–176
190. Christianto V, Smarandache F (2019) A review of seven applications of neutrosophic logic: in cultural psychology, economics theorizing, conflict resolution, philosophy of science, etc. J — Multidiscip Sci J 2(2):128–137
191. Devadoss AV, Anand MCJ, Felix A (2012) A study on the impact of violent video-games playing among children in Chennai using neutrosophic cognitive maps (NCMs). Int J Sci Eng Res 3(8):1–4
192. Yasser I, Twakol A, El-Khalek AAA, Samrah A, Salama AA (2020) COVID-X: Novel health-fog framework based on neutrosophic classifier for confrontation Covid-19. Neutrosophic Sets Syst 35:1–21

193. Salama AA, El-Ghareeb HA, Mani AM, Lotfy MM (2014) Utilizing neutrosophic set in social network analysis e-learning systems. Int J Inf Sci Intell Syst 3(2):1–12
194. Radwan NM (2016) Neutrosophic applications in E-learning: outcomes, challenges and trends. In: Smarandache F, Pramanik S (eds) New trends in neutrosophic theory and application. Pons Editions, pp 177–184
195. Anitha R, Gunavathi K (2017) NCM-based raga classification using musical features. Int J Fuzzy Syst 19(5):1603–1616
196. Shadrach FD, Kandasamy G (2020) Neutrosophic cognitive maps (NCM) based feature selection approach for early leaf disease diagnosis. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02070-3
197. Pamucar D, Deveci M, Schitea D, Erişkin L, Iordache M, Iordache I (2020) Developing a novel fuzzy neutrosophic numbers based decision making analysis for prioritizing the energy storage technologies. Int J Hydrogen Energy. https://doi.org/10.1016/j.ijhydene.2020.06.016
198. Albeanu G (2013) Neutrosophic computational models–(I). Ann Spiru Haret Univ Math-Inf Ser 9(2):13–22
199. Bencze M (2006) Neutrosophic applications in finance, economics and politics a continuing bequest of knowledge by an exemplarily innovative mind. Scientia Magna 2(4):81–82
200. Bhattacharya S (2005) Utility, rationality and beyond-from behavioral finance to informational finance. American Research Press
201. Bhattacharya S, Smarandache F, Kumar K (2005) Conditional probability of actually detecting a financial fraud-a neutrosophic extension to Benford's law. Int J Appl Math 17 (1):7–14
202. Khoshnevisan M, Bhattacharya S (2002) A short note on financial data set detection using neutrosophic probability. In: Proceedings of the first international conference on Neutrosophy, neutrosophic logic, neutrosophic set, neutrosophic probability and statistics. Phoenix, Xiquan, pp 75–80
203. Khoshnevisan M, Bhattacharya S (2005) Neutrosophic information fusion applied to the options market. Investment Manage Fin Innov 1:139–145
204. Bhattacharya S, Khoshnevisan M, Smarandache F (2003) Artificial intelligence and responsive optimization. Xiquan, Phoenix
205. Bhattacharya S, Smarandache F, Khoshnevisan M (2006) The Israel-Palestine question—a case for application of neutrosophic game theory. Octogon Math Maga 14(1):165–173
206. Plessner Y (2001, February 15) The conflict between Israel and the Palestinians: a rational analysis. Jerusalem Letters/Viewpoints, No 448, 22 Shvat 5761
207. Pramanik S, Roy TK (2013) Game theoretic model to the Jammu-Kashmir conflict between India and Pakistan. Int J Math Archive 4(8):162–170
208. Pramanik S, Roy TK (2014) Neutrosophic game theoretic approach to Indo-Pak conflict over Jammu-Kashmir. Neutrosophic Sets Syst 2:82–101
209. Deli I (2019) Matrix games with simplified neutrosophic payoffs. In: Fuzzy multi-criteria decision-making using neutrosophic sets. Springer, Berlin, pp 233–246
210. Vasantha WB, Kandasamy I, Smarandache F, Devvrat V, Ghildiyal S (2020) Study of imaginative play in children using single-valued refined neutrosophic sets. Symmetry 12 (3):402
211. Kandasamy I, Vasantha WB, Obbineni JM, Smarandache F (2020) Sentiment analysis of tweets using refined neutrosophic sets. Comput Ind 115:103180. https://doi.org/10.1016/j.compind.2019.103180
212. Vasantha WB, Kandasamy I, Devvrat V, Ghildiyal S (2019) Study of imaginative play in children using neutrosophic cognitive maps model. Neutrosophic Sets Syst 30:241–252
213. Devadoss AV, Rekha M (2017) A study on child marriage using new neutrosophic associative fuzzy cognitive dynamical system. Global J Pure Appl Math 13(6):2255–2265
214. Mondal K, Pramanik S (2014) A study on problems of Hijras in West Bengal based on neutrosophic cognitive maps. Neutrosophic Sets Syst 5:21–26

215. Pramanik S, Chackrabarti SN (2013) A study on problems of construction workers in West Bengal based on neutrosophic cognitive maps. Int J Innov Res Sci Eng Technol 2(11):6387–6394
216. Jousselme AL, Maupin P (2004) Neutrosophy in situation analysis. In: Proceedings of fusion, pp 400–406
217. Thiruppathi P, Saivaraju N, Ravichandran KS (2010) A study on suicide problem using combined overlap block neutrosophic cognitive maps. Int J Algorithms Comput Math 3 (4):22–28
218. Zafar A, Anas M (2019) Neutrosophic cognitive maps for situation analysis. Neutrosophic Sets Syst 29:78–88
219. William MA, Devadoss AV, Sheeba JN (2013) study on neutrosophic cognitive maps (NCMs) by analyzing the risk factors of breast cancer. Int J Sci Eng Res 4(2):1–4
220. Kandasamy WBV, Smarandache F (2004) Analysis of social aspects of migrant laborers living with HIV/AIDS using fuzzy theory and neutrosophic cognitive maps. Xiquan, Phoenix
221. Kandasamy V, Smarandache, F (2005) Fuzzy neutrosophic analysis of women with HIV/AIDS. Hexis, Arizona
222. Bernajee G (2008) Adaptive fuzzy cognitive maps vs neutrosophic cognitive maps: decision support tool for knowledge based institution. J Sci Ind Res 67:665–673
223. Ramalingam S, Kandasamy WB, Broumi S (2019) An approach for study of traffic congestion problem using fuzzy cognitive maps and neutrosophic cognitive maps-the case of Indian traffic. Neutrosophic Sets Syst 30:273–283
224. Saha A, Smarandache F, Baidya J, Dutta D (2020) MADM using m-generalized q-neutrosophic sets. Neutrosophic Sets Syst 35:252–268
225. Das S, Pramanik S (2020) Generalized neutrosophic b-open sets in neutrosophic topological space. Neutrosophic Sets Syst 33:552–530
226. Salama AA, Smarandache F (2015) Neutrosophic crisp set theory. The Educational Publisher Inc., Columbus. http://fs.unm.edu/NeutrosophicCrispSetTheory.pdf
227. Thong NT, Dat LQ, Hoa ND, Ali M, Smarandache F (2019) Dynamic interval valued neutrosophic set: modeling decision making in dynamic environments. Comput Ind 108:45–52
228. Kandasamy I (2018) Double-valued neutrosophic sets, their minimum spanning trees, and clustering algorithm. J Intell Syst 27(2):163–182
229. Karaaslan F, Hunu F (2020) Type-2 single-valued neutrosophic sets and their applications in multi-criteria group decision making based on TOPSIS method. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-01686-9
230. Broumi S, Smarandache F (2014) Single valued neutrosophic trapezoid linguistic aggregation operators based multi-attribute decision making. Bull Pure Appl Sci- Math Stat 33(2):135–155
231. Deli I, Şubaş Y (2017) Some weighted geometric operators with SVTrN-numbers and their application to multi-criteria decision making problems. J Intell Fuzzy Syst 32(1):291–301
232. Şahin R, Liu P (2017) Possibility-induced simplified neutrosophic aggregation operators and their application to multi-criteria group decision-making. J Exp Theor Artif Intell 29(4):769–785
233. Deli I (2017) Interval-valued neutrosophic soft sets and its decision making. Int J Mach Learn Cybern 8(2):665–676
234. Broumi S (2013) Generalized neutrosophic soft set. Int J Comput Sci Eng Inf Technol 3 (2):17–30
235. Broumi S, Sahin R, Smarandache F (2014) Generalized interval neutrosophic soft set and its decision making problem. J New Res Sci 7:29–47
236. Alkhazaleh S (2016) Time-neutrosophic soft set and its applications. J Intell Fuzzy Syst 30 (2):1087–1098

237. Ulucay V, Şahin M, Olgun N (2018) Time-neutrosophic soft expert sets and its decision making problem. MATEMATIKA: Malaysian J Industrial Appl Math 34(2):246–260
238. Broumi S, Deli I, Smarandache F (2014) Interval valued neutrosophic parameterized soft set theory and its decision making. J New Res Sci 3(7):58–71
239. Chatterjee R, Majumdar P, Samanta SK (2016) Interval-valued possibility quadripartitioned single valued neutrosophic soft sets and some uncertainty based measures on them. Neutrosophic Sets Syst 14:35–43
240. Zhao AW, Guan HJ (2015) Neutrosophic valued linguistic soft sets and multi-attribute decision-making application. J Comput Theor Nanosci 12(12):6162–6171
241. Deli I; Şubaș Y, Smarandache F, Ali M (2016) Interval valued bipolar neutrosophic sets and their application in pattern recognition. https://arxiv.org/abs/1601.01266. Accessed on 9 July 2020
242. Deli I, Şubaş Y, Smarandache F, Ali M (2016) Interval valued bipolar fuzzy weighted neutrosophic sets and their application. In: 2016 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, New York, pp 2460–2467
243. Ali M, Smarandache F (2017) Complex neutrosophic set. Neural Comput Appl 28:1817–1834
244. Ali M, Dat LQ, Smarandache F (2018) Interval complex neutrosophic set: formulation and applications in decision-making. Int J Fuzzy Syst 20(3):986–999
245. Al-Quran A, Hassan N (2018) The complex neutrosophic soft expert set and its application in decision making. J Intell Fuzzy Syst 34(1):569–582
246. Dat LQ, Thong NT, Ali M, Smarandache F, Abdel-Basset M, Long HV (2019) Linguistic approaches to interval complex neutrosophic sets in decision making. IEEE Access 7:38902–38917
247. Broumi S, Bakali A, Talea M, Smarandache F, Singh PK, Uluçay V, Khan M (2019) Bipolar complex neutrosophic sets and its application in decision making problem. In: Fuzzy multi-criteria decision-making using neutrosophic sets. Springer, Berlin, pp 677–710
248. Cruz RA, Irudayam FN (2017) Neutrosophic soft cubic set. Int J Math Trends Technols 46 (2):88–94
249. Xue H, Yang X, Chen C (2020) Possibility neutrosophic cubic sets and their application to multiple attribute decision making. Symmetry 12(2):269
250. Broumi S, Smarandache F (2015) Interval-valued neutrosophic soft rough sets. Int J Comput Math Article Id:232919. https://doi.org/10.1155/2015/232919
251. Ye J (2016) Correlation coefficients of interval neutrosophic hesitant fuzzy sets and its application in a multiple attribute decision making method. Informatica 27(1):179–202
252. Fahmi A, Aslam M, Riaz M (2020) New approach of triangular neutrosophic cubic linguistic hesitant fuzzy aggregation operators. Granular Comput 5:527–543
253. Awang A, Ali M, Abdullah L (2019) Hesitant bipolar-valued neutrosophic set: formulation, theory and application. IEEE Access 7:176099–176114
254. Ye S, Ye J (2014) Dice similarity measure between single valued neutrosophic multisets and its application in medical diagnosis. Neutrosophic Sets Syst 6:49–54
255. Kandasamy I, Smarandache F (2016) Triple refined indeterminate neutrosophic sets for personality classification. In: 2016 IEEE symposium series on computational intelligence (SSCI). https://doi.org/10.1109/ssci.2016.7850153
256. Kandasamy I, Smarandache F (2017) Multicriteria decision making using double refined indeterminacy neutrosophic cross entropy and indeterminacy based cross entropy. Appl Mech Mater 859:129–143
257. Smarandache F (2019) Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, q-rung orthopair fuzzy set, spherical fuzzy set, and n-hyperspherical fuzzy set, while neutrosophication is a generalization of regret theory, grey system theory, and three-ways decision (revisited). https://arxiv.org/abs/1911.07333

258. Das S, Roy BK, Kar MB, Kar S, Pamučar D (2020) Neutrosophic fuzzy set and its application in decision making. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-01808-3
259. Alkhazaleh S (2015) Neutrosophic vague set theory. Crit Rev 10:29–39
260. Hashim H, Abdullah L, Al-Quran A (2019) Interval neutrosophic vague sets. Neutrosophic Sets Syst 25:66–75
261. Al-Quran A, Hassan N (2017) Neutrosophic vague soft set and its applications. Malaysian J Math Sci 11(2):141–163
262. Al-Quran A, Hassan N (2016) Neutrosophic vague soft expert set theory. J Intell Fuzzy Syst 30(6):3691–3702
263. Mukherjee A (2019) Vague–valued possibility neutrosophic vague soft expert set theory and its applications. Neutrosophic Sets Syst 29:142–157
264. Al-Quran A, Hassan N (2018) Neutrosophic vague soft multiset for decision under uncertainty. Songklanakarin J Sci Technol 40(2)
265. Hassan N, Al-Quran A (2017) Possibility neutrosophic vague soft expert set for decision under uncertainty. In: AIP conference proceedings, vol 1830, No 1. AIP Publishing LLC, p. 070007
266. Hussain SS, Hussai RJ, Jun YB, Smarandache F (2019) Neutrosophic bipolar vague set and its application to neutrosophic bipolar vague graphs. Neutrosophic Sets Syst 28:69–86
267. Bhowmik M, Pal M (2009) Intuitionistic neutrosophic set. J Inf Comput Sci 4(2):142–152
268. Broumi S, Smarandache F (2013) Intuitionistic neutrosophic soft set. J Inf Comput Sci 8 (2):130–140
269. Broumi S, Deli I, Smarandache F (2015) N-valued interval neutrosophic sets and their application in medical diagnosis. Crit Rev 10:45–69
270. Peng HG, Zhang HY, Wang JQ (2018) Probability multi-valued neutrosophic sets and its application in multi-criteria group decision-making problems. Neural Comput Appl 30 (2):563–583
271. Wang N, Zhang H (2017) Probability multivalued linguistic neutrosophic sets for multi-criteria group decision-making. Int J Uncertain Quantif 7(3):207–228
272. Liu P, Teng F (2018) Multiple attribute decision making method based on normal neutrosophic generalized weighted power averaging operator. Int J Mach Learn Cybern 9:281–293
273. Kamal NM, Abdullah L (2019) Multi-valued neutrosophic soft set. Malaysian J Math Sci 13:153–168
274. Kamal NLAM, Abdullah L, Abdullah I, Alkhazaleh S, Karaaslan F (2019) Multi-valued interval neutrosophic soft set: formulation and theory. Neutrosophic Sets Syst 30:149–170
275. Alkhazaleh S (2017) n-Valued refined neutrosophic soft set theory. J Intell Fuzzy Syst 32 (6):4311–4318
276. Jha S, Kumar R, Chatterjee JM, Khari M, Yadav N, Smarandache F (2019) Neutrosophic soft set decision making for stock trending analysis. Evol Syst 10(4):621–627
277. Tey DJY, Gan, YF, Selvachandran G, Quek SG, Smarandache F, Son LH, Long HV (2019) A novel neutrosophic data analytic hierarchy process for multi-criteria decision making method: a case study in Kuala Lumpur Stock exchange. IEEE Access 7:53687–53697
278. Bustince H, Burillo P (1996) Vague sets are intuitionistic fuzzy sets. Fuzzy Sets Syst 79:403–405
279. Deschrijver G, Kerre EE (2003) On the relationship between some extensions of fuzzy set theory. Fuzzy Sets Syst 133(2):227–235

Surapati Pramanik has been working as a teacher educator since 2006 in Nandalal Ghosh B.T. College, India. He received a Ph.D. in Mathematics from Indian Institute of Engineering Science and Technology (IIEST), Shibpur, Ph.D. in Education, an M.Sc., and an M.Ed. from University of Kalyani. He acts as a Ph.D. supervisor in Mathematics for Jadavpur University and IIEST, Shibpur. His research focuses on fuzzy multi-criteria decision-making (MCDM), mathematics education, soft computing, and neutrosophic fuzzy sets. His research earned outstanding paper awards several times in West Bengal State Science and Technology Congress in mathematics (2011) and Social Sciences (2010, 2013, 2019). He is an editorial board member of "Neutrosophic Sets and Systems", and "Current Chinese Science: Artificial Intelligence & Robotics". His publication includes 24 book chapters, 130 research papers, and two editorial books. He acts as a reviewer for more than 40 international journals with more than 100 reviews. He is a life member of thirteen scientific organizations including Calcutta Mathematical Society, and ISI, Kolkata. Currently, he is working on neutrosophic and fuzzy hybrid sets.

27 Technology Assessment Integrated Research for Scientific Policy Advice

Armin Grunwald

In the years ahead, as we move into the twenty-first century, there will be even greater need to rely on … Technology Assessment for impartial assessment of technology-related policies. The world of science and its impact on public policy are becoming more complex, not less. Technology is central to every aspect of … life, from biotechnology to law enforcement, from agriculture to education.

Senator Edward Kennedy, speech delivered to the U.S. Senate in July 1995

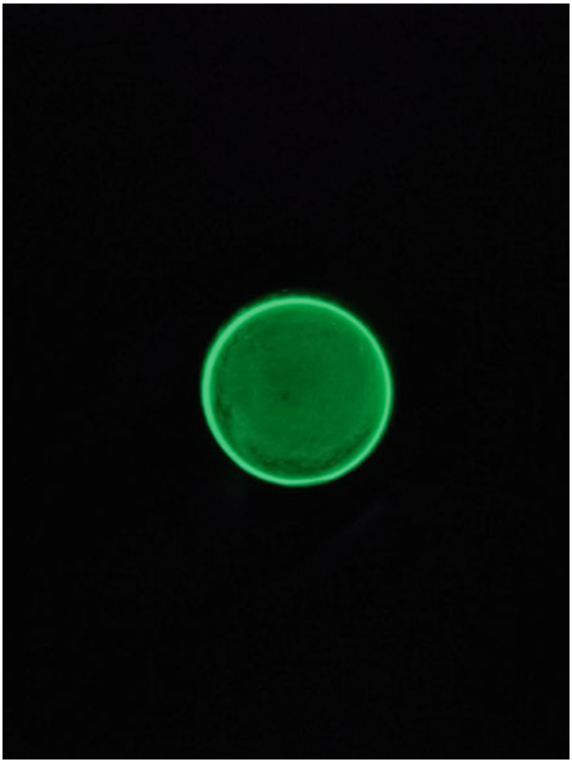
Summary

Technology assessment (TA) constitutes an integrative research field providing knowledge and orientation for better-informed and well-reflected decisions concerning new technologies. Amongst its motivation is exploring answers in case of orientation demands or conflicts around science and technology, e.g. at the occasion of the occurrence of unintended side effects such as risks, or for providing strategies of responsible innovation. The consequences of new technology do not stop at the boundaries of scientific disciplines. Rather, they usually touch upon technological, ecological, economic, social, and ethical issues, amongst others. Therefore, technology assessment applies approaches to and methods of knowledge integration across disciplines. Beyond this interdisciplinary character, technology assessment also has to involve knowledge, values, and perspectives of persons and groups beyond science, such as citizens and stakeholders. Hence, technology assessment often also works in a

A. Grunwald (✉)
Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Karlstr. 11, 76133 Karlsruhe, Germany
e-mail: armin.grunwald@kit.edu

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_27

transdisciplinary manner. Its aim always is supporting the evaluation of technologies and their impact according to societal values and ethical principles, elaborating strategies to deal with the uncertainties that inevitably arise, and contributing to constructive solutions of societal conflicts around the development, use, and disposal of new technologies. The chapter presents an overview of technology assessment status reached so far, thereby focusing on issues of integration.



[Adapted with permission from the Association of Science and Art (ASA), USERN; Artist: Sepideh Sargoli].

The code of this chapter is 01010100 01000001.

1 Introduction and Overview

Technology Assessment (TA) has been emerging since the 1960ies. It constitutes an interdisciplinary research field aiming at providing knowledge for better-informed and well-reflected decisions concerning new technologies [1]. TA shall enhance the reflexivity of deliberating about new technology, innovation, and their societal consequences. Exploring, assessing, and evaluating consequences of new technology needs integrating any available knowledge on possible side effects at an early stage in decision-making processes, evaluating of technologies and their impact according to societal values and ethical principles, elaborating strategies to deal with the uncertainties that inevitably arise, and contributing to constructive solutions of societal conflicts [1]. In this sense, technology assessment may be regarded element of scientific approaches to enable and empower responsible advance in the framework of the 'risk society' [2] and the reflexive modernization [3, 4].

The consequences of new technology do not stop at the boundaries of scientific disciplines. Rather, they usually touch upon technological, ecological, economic, social, political, cultural, and ethical issues, amongst possibly others. Therefore, technology assessment applies approaches to and methods of knowledge integration across disciplines. Beyond this *interdisciplinary* character, technology assessment often also has to integrate knowledge, values, and perspectives of persons and groups beyond science, such as citizens and stakeholders. Hence, TA frequently also operates in a *transdisciplinary* manner. In the past five decades, technology assessment has gained huge experience in these fields of integration, e.g. by providing integrated and coherent advice to many parliaments, by informing public debate and by integrating its advice directly into the engineering and design of new technologies.

The chapter presents an overview of the status of technology assessment reached so far according to [1, 5, 6], focused on the issue of integration and neighboured areas specific to the field of TA. Large parts of the description also apply to the closely related field of responsible research and innovation [7–9]. Based on a brief presentation of the major motivations of TA and its historical background (Sect. 2), the TA framework will be introduced (Sect. 3), which allows giving a systematic introduction into the needs for integrating knowledge (Sect. 4), and presenting a procedural approach to integration in TA (Sect. 5).

2 Technology Assessment: Motivations and History

Since the 1960s, the disadvantages of scientific and technical innovations became obvious some of them were of dramatic proportions. Accidents in technical facilities (Chernobyl, Bhopal), natural environment's threats (air and water pollution, climate change, loss of biodiversity), negative health effects as in the asbestos case, social and cultural side effects (e.g. labour market problems due to automation) and the intentional abuse of technology (e.g. the attacks on the World Trade Centre in 2001) are well known examples. Parallel to these developments, broad segments of Western society were confronted with scenarios of 'Limits to Growth' [10], which, for the first time, addressed the grave resource problems perceived as a side effect of technology and economic growth. Deepened insight into technological ambivalence led to a crisis of orientation in the way society should deal with science and technology in its further development. This still ongoing and probably persistent crisis forms the most essential background of and reason for the emergence of TA. The huge impact of today's technology in the *Anthropocene* [11] and the spreading of many of its consequences across space and time demonstrate that introducing new technology by following a 'trial and error' approach or a 'wait and see' strategy cannot be considered responsible. By far not the only one but probably the most pressing questions fuelling TA, therefore, are: How can a society, which requires to trust the innovation, protect itself from undesirable side effects? How can humankind benefit from the technological advance in a responsible and sustainable manner? How can society preventatively act to cope with possible future adverse effects? In order to enable managing the emergence of possible unintended side effects and the scarcity of resources, regulatory and scientific approaches were developed with technology assessment being one of the major approaches based on science [1]. Their overall mission is to extend foresight and reflection capabilities in the huge field of scientific and technological advances and for making the best use of its outcomes.

Additional motivations entered the field of technology assessment over the past decades. They enriched the manifold of expectations and activity fields of TA and motivated several focal points, thereby involving specific methodologies and conceptual debates. The most significant extensions are (following [1], Chap. 4):

- *Shaping technology according to social values*: In the framework of social constructivism [12], the approach was born to design technology according to social values. If this would succeed, so the hope, unintended side effects could partially be avoided. Instead, technology assessment was dedicated to enabling a 'better technology in a better society' [5]. In this respect, the framework of sustainable development received major attention for guiding the assessments [13].
- *Combatting technocracy and empowering democracy*: in the 1960s, concerns emerged that the scientific and technological advance could threaten the functioning of democracy [14]. The technocracy hypothesis painted a picture of a future society. Experts, according to this concern, would take over control. Democratically legitimated policymakers could not prevent this due to a lack of knowledge. TA should, and still shall, enable and empower society to take active roles in democratic deliberation on science and technology [15, 16].
- *Contributing to conflict mediation*: little acceptance of some political decisions on technology such as on nuclear power in some countries as well as doubts about their legitimacy and resulting conflicts motivated TA to think about procedures of conflict prevention and resolution, in particular including participatory approaches for broadening democratic legitimization [17].
- *Understanding techno-visionary communication*: political and public debates on techno-visionary sciences such as nanotechnology, enhancement technologies, synthetic biology, and artificial intelligence have been accompanied by far-ranging futurist communication. The widespread use of visions, utopian and dystopian narratives, and metaphors became an important factor challenging traditional TA [8, 18].

In modern TA, it is often not only a question of the consequences of individual technologies, products, or plants, but frequently of complex configurations between enabling technologies, innovation potentials, fears and concerns, patterns of production and consumption, lifestyle and culture, and political and strategic decisions [6]. TA usually does not consider technology and its consequences *as such* but rather technology embedded in social practices such as the economy, lifestyle, governmental action, and public communication. Therefore, TA has to consider future *socio-technical* configurations and the co-evolution of technology and society [19], rather than the consequences of technology alone. This happens in specific arenas consisting of different actors, institutional settings, expectations, and positions in the overall technology governance. They can be classified into three fields, according to [1]:

1. *TA as policy advice*: In this field, the objective of TA is to support policymakers in addressing the above-mentioned challenges (e.g. [20]), e.g. by providing them with foresight information, ethical reflection, and alternative strategies on how to proceed in the respective field. Here, TA does not *directly* address technology development but considers the *boundary conditions* of technology development and use. *Parliamentary* TA is a sub-category of policy-advising TA presupposing that parliaments play a crucial or at least an important and relevant role in technology governance [21, 22].
2. *TA as a contribution to public dialogue*: participatory TA developed approaches to involve citizens, consumers and users, actors of civil society, stakeholders, the media and the public in different roles at different stages in technology governance [17, 23]. According to normative ideas of deliberative democracy, the assessment of technology should be left neither to the scientific experts (expertocracy) nor to the political deciders alone (decisionism, cp. [14]). Several approaches and methods have been developed and applied in recent decades, such as consensus conferences, citizens' juries, and focus groups [23].
3. *TA as part of technology development*: Building on research on the genesis of technology made in the framework of social constructivism [12], the approach of *shaping technology* due to social expectations and values motivated the development of several approaches such as Constructive TA (CTA, cp. [5]). TA in this respect aims at enriching technology development and engineering by foresight and reflection, thereby addressing the level of concrete products, systems, and services.

3 Conceptual Framework of Technology Assessment[1]

The categorization given above shows a broad variety of obviously heterogeneous TA practices covering different actor constellations, involving different role concepts for TA, and responding to different expectations by applying different concepts and methods. In order to identify a common TA framework for subsuming the many and various TA activities the crucial step in determining the overall *cognitive*

[1] This and the following section include results from former publications, e.g. from [1], Chap. 4, in a dense and summarized form.

interest of technology assessment. A broad consideration of the fields of TA practice allows for identifying the common cognitive interest of TA as [1, p. 88].

> supporting, strengthening and enhancing reflexivity in all epistemic and social fields of reasoning and decision-making on shaping the scientific and technological advance, on the usage of its outcomes and on dealing with the consequences to present and future society.

However, the notion of enhancing reflexivity is rather abstract. In order to make it more tangible, three conceptual dimensions of enhancing reflexivity bridging the gap between TA's cognitive interest mentioned above and the fields of TA practice (previous section) have been identified (Fig. 1; cp. Sect. 4 of this chapter):

- in the dimension of *anticipation* (e.g. [24, 25]), TA aims at enhancing reflexivity *over time* by consideration of broader future developments, in particular by systematically taking into account the possible occurrence of unintended effects [6, 26]
- in the dimension of *inclusion,* TA aims at enhancing reflexivity *across perspectives* by involving a broader range of actors such as stakeholders and citizens with their knowledge, values, and interests (e.g. [17, 23, 27])
- in the dimension of *complexity* TA aims at enhancing reflexivity *over relevance* by critically scrutinizing the judgements on relevance involved, e.g. by determining system boundaries.

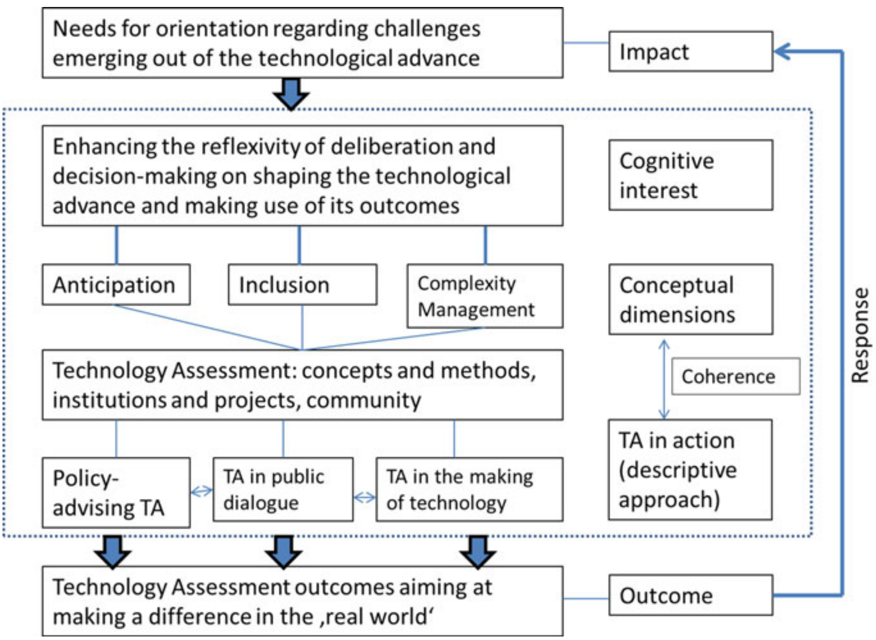


Fig. 1 Framework of TA (*Source* [1], modified)

Figure 1 provides an overall picture of TA according to these dimensions (cp. [1] for a more detailed derivation and explanation). At its top, the societal needs and demands for orientation are mentioned. TA, working in the dotted box, produces outcomes as responses to those needs and demands (at the bottom), which shall create an impact in the real world. In this way, TA is part of research-based, societal learning processes [28] for dealing with technological advancements and their outcomes in a reflexive manner. TA shall simultaneously conduct and enlighten these processes beyond unrecognized and non-reflected pre-occupations and ideology. Therefore, ensuring transparency and argumentative clarity in normative as well as in epistemological respect are strong obligations to TA.

Enhancing reflexivity in the dimensions mentioned is obviously a necessity in all TA branches mentioned above: in supporting democratic reasoning by policy advice, in strengthening public debate as well as directly enriching the making of technology. The necessity to avoid 'wait and see' or 'trial and error' approaches in the Anthropocene ([11], cp. also the 'Imperative of Responsibility [29]) concerns the policymaking processes, public debate, and responsible engineering as well. Increasing human power through science and technology implies increased human responsibility. Enhancing reflexivity by involving technology assessment aims at making this feasible.

Several needs for integration can easily be derived from this conceptual scheme. At the top, this need is rooted in practical demands for orientation. Policymakers and societal actors would not be satisfied to receive from TA a mixture of different and heterogeneous, perhaps inconsistent pieces of knowledge. Instead, they demand integrated perspectives on the technology under consideration covering a broad range of knowledge types, e.g. from engineering, social science, and ethics, as well as a balanced view on normative issues such as values and interests involved. In particular, this holds for policymakers because they always have to make decisions, e.g. on regulation or promotion of new technology, based on large sets of sub-questions to be considered together. The fact that decisions to be made have to integrate many different issues is at the core of the integration necessity in TA. This includes, obviously, the integration of knowledge from various disciplinary fields of science as well as the integration of ethical values and social perspectives.

4 Inclusion and Knowledge Integration

In this section, I will deepen the consideration of integration in technology assessment. To this end, I refer to the conceptual dimensions of TA briefly introduced above: anticipation, inclusion, and complexity (following [1], Chap. 4).

4.1 Anticipation

TA is in need of *anticipation* of possible futures of society and human life involving possible future science and technology [29–32] because otherwise it would not be possible to provide support to opinion-forming and decision-making for shaping issues of the future. TA creates and assesses *prospective knowledge* about possible future consequences of technology, be they intended or unintended, as a basis for drawing conclusions for today's decision-making (Fig. 2). In this mode of operation, it works under the *consequentialist paradigm*. In the German translation of technology assessment, *Technikfolgenabschätzung,* the notion of consequences (*Folgen*), has even been made part of the term. Prospective knowledge in TA addresses consequences, which do not yet exist and perhaps will never become a reality [33]. Anticipation serves the purpose of enhancing reflexivity with respect to future developments in order to improve the basis for decision-making. It covers a broad range of various types of images of the future. Model-based simulations, trend extrapolations, scenarios about how a future world would or could look if the respective technology was developed, produced, and implemented, but also speculative visions, normative hopes and expectations, fears and even utopias and dystopias have to reflected in TA in different modes of creating orientation [34]. Thereby, anticipatory, integrated and coherent prospective knowledge shall become a catalyst for deliberating alternative futures *in the present* for shaping technology and its social environment such as regulation and incentive measures *in some future*. The major expectation is that TA shall contribute to better-reflected decisions after having performed the consequentialist loop (Fig. 2).

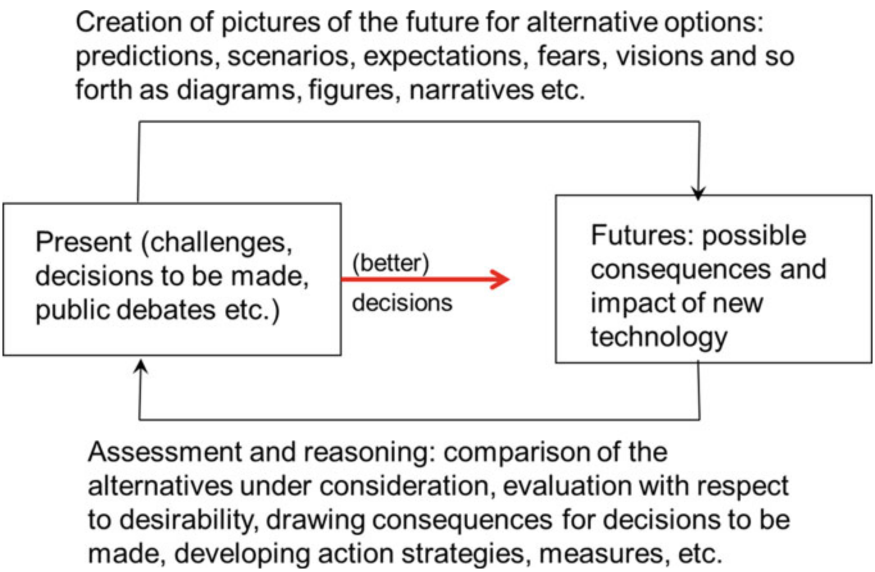


Fig. 2 TA in the consequentialist paradigm (*Source* [35], p. 313, modified)

4.2 Inclusion

Technology assessment has to involve and include perspectives and values of different actors and groups affected by the technology under consideration [17, 36]. It needs in particular to consider the possible consequences of actions and decisions not only from the perspective of decision-makers. Rather, TA takes into account the possibly diverse and diverging perspectives and different pieces of knowledge supplied by stakeholders, local authorities, citizens, and people affected by those consequences, including future generations. Hence, inclusion in TA means that the diversity of values and standpoints must be taken into account in the assessment, as well as knowledge contributions submitted by the different social actors involved. Inclusion is a *social* process with legitimization-creating properties as well as an *epistemic* process integrating additional knowledge. Inclusion is, therefore, confronted by the challenge to integrate varieties of different kinds: knowledge with different and often incommensurable epistemic properties, values and moral convictions, which can contradict each other, interests and degrees of affectedness by the technologies under consideration, and so forth. Integration in TA, thus, extends beyond knowledge integration but also concerns the integration of normative entities such as values.

The latter point involves a specific background of modern society. Individualization in modern times has immediate consequences for how to assess science and technology and their outcomes. Individual assessments will follow diverse and individual values, interests and convictions and will usually arrive at subjective results. In particular, technologies closely related to morally sensitive issues in medicine often create challenges in this respect: at the beginning and end of human life, abortion, transplantation technologies or human enhancement technologies touching questions of the value of human life. In these fields but also for dealing with issues of risk induced by technology various and often diverging perceptions of benefits and risks are distributed across individuals and groups. Frequently, those persons or groups benefitting from specific decisions are different from the groups and persons possibly burdened by unintended effects, e.g. noise or other emissions. Therefore, it is crucial for integrative assessment processes that issues of distribution are dealt with in a fair and just manner [27].

The major expectation concerning inclusion is that orientation with broad support and appreciation in wide parts of society will be more robust and better legitimated, e.g. with respect to changes of political majorities or in cases of conflict (e.g. [37]). In epistemic respect, the inclusion of knowledge from different sources also from out of the scientific disciplines will support mobilizing the best arguments and contribute to the better durability of the resulting orientation. Again, practical requirements and desires lead to the necessity of integration.

4.3 Complexity

Enhancing reflexivity for anticipation and inclusion as described above is at the core of TA but must not and cannot be an *absolute* category. Something still is missing: a stop-rule when enhancing reflexivity comes to an end. Enhancing reflexivity in anticipation and inclusion does not have a built-in stop-rule: we can consider more and more dimensions of possible consequences without observing their plausibility, and we can include more and more people and perspectives. However, it is neither possible nor necessary to include everything and everyone, and similarly, it is neither possible nor necessary to consider every thinkable future narrative. Complexity is overwhelming anyway and can explode without applying some stop-rule. What is the stop-rule for avoiding complexity explosion, which would destroy any TA project? The key for determining stop-rules are considerations of *relevance*.

TA aiming at exploring and assessing new technology by looking at its consequences in an inclusive manner often conceptualizes the technology, its social environment and the range of consequences *as a system* with boundaries, with an external environment and internal elements, with interdependencies, and with input/output relations at the system boundaries. This conceptualization allows applying a familiar scheme of research: identification of the elements of the system, investigation of their interrelations, research into the input/output-relations, sensitivity analyses, and so forth. Determining a system means drawing a line between the included and the excluded, which needs *reflection on what is regarded relevant, and why*.

Reflections on relevance are particularly important because systems do not exist *as such.* Rather, arrangements of the real or imagined world are constituted *as systems* with respect to specific research or orientation interests. Determining system boundaries in temporal, spatial, thematic, and social respects is a crucial activity in TA, e.g. in the field of sustainability and life cycle assessments. Constituting something as a system that excludes the world outside from closer consideration which simplifies the exercise. Usually, the exchange of the system under consideration with the external world is taken into account by addressing *input/output relations* between the system and its environment. Often, a system in TA is defined in a way that needs distinct input from its environment and gives back (intended) services but possibly also (unintended) risks and costs. With regard to this framework the performance—the relations between input and output—of different systems involving different technological solutions can be compared (cost/benefit ratio, efficiency, emissions, resource productivity …). In order to summarize: system thinking and the use of methods taken from systems analysis belong to the crucial conceptual dimensions of TA, because they trigger reflections on relevance at each stage of the assessment process.

Integration is an important issue also here because decisions on relevance and the subsequent determination of adequate system boundaries have to take into account different pieces of knowledge but also normative criteria of why something is regarded relevant to the problem under consideration. As an example, sensitivity analyses often are based on integrative models bringing together knowledge from different disciplines, e.g. in climate research and in sustainability science.

5 Assessment as Procedural Integration

The concurrent involvement of *epistemic* elements—the pieces of knowledge involved—and *social* issues of inclusion and legitimization allows denoting TA as a set of *socio-epistemic practices* [1]. At the first glance, the allusion of social and epistemic issues could create the impression of compromising scientific rationality. However, as Habermas [38] pointed out, inclusive democratic deliberation and scientific reasoning are commonly rooted in argumentative rationality. Hence, the gap between the two seemingly completely different areas of inclusive democratic deliberation and decision-making on the one hand, and scientific reasoning on the other, can be bridged by a strong dedication to transparency and argumentative rationality. At the core of TA is the argument-based search for good solutions in a broadened range of futures (anticipation) and actors (inclusion), governed by stop-rules based on relevance considerations. This constellation demonstrates that the demand for integration in TA exceeds usual postulates of inter—and transdisciplinary integration of knowledge. It also extends to arranging the integrative assessment in a specific manner dedicated and obliged to requirements of sound epistemology and social legitimacy as well. The complexity of integration in TA can be nicely illustrated by taking a look to the various types of knowledge and orientation to be integrated (Table 1; cp. [1]).

Integration in TA involves necessities in two respects, which are particularly challenging in the assessment process: (1) integrating pieces of knowledge from different disciplinary and extra-scientific sources, which may have incommensurable claims for validity, and (2) integrating assessment and evaluation results according to different thematic dimensions and criteria into an overarching picture. Knowledge types and pieces created, used, and assessed in TA come from different sources and are generally rather heterogeneous and in part incommensurable in epistemic respects. Natural sciences such as geography, ecology and climatology, as well as social sciences, provide systems knowledge. Disciplines such as ethics and legal sciences contribute to normative orientation. Future studies provide the assessment with scenarios, and hermeneutic inquiry helps in understanding the respective context, whilst action-oriented sciences such as politics and economics deal with measures to reach specific targets.

Integration is often epistemologically difficult and sometimes seems even to be impossible. The decisive criterion is the *epistemic interdependence* of different pieces of knowledge. As long as these pieces do not interfere in epistemic respects, i.e. if there is no overlap or mutual dependency according to their premises and underlying theories, integration is easy and simply consists of knowledge addition. For example, there is probably no epistemic interference between geological

Table 1 Classification of knowledge involved in technology assessment (*Source* [1])

Knowledge in TA	Functions	Attributes	Objects	Disciplines
Systems knowledge	Understanding the functionality of the system and considering its boundaries	General, descriptive and empirics-based in the form of causal or statistical knowledge	Empirical natural and social systems, socio-technical constellations	Natural, social, and engineering sciences, Earth systems analysis, STS studies (…)
Prospective knowledge	Illustrating the space of possible or plausible future developments for enabling consequentialist reflection	Explorative and extrapolative projections based on present systems knowledge involving different assumptions	Development of specific parameters or indicators in the future	Futures studies in different fields, simulation, foresight, systems analysis (…)
Normative orientation	Provide criteria of evaluation and assessment and targets of transformation	Normative, based on ethical or legal reasoning	Action and decision-making, weighing up alternative options, trade-offs	Ethics, legal sciences, political theory (…)
Hermeneutic knowledge	Understanding the specific case and its social as well as epistemic configuration	Interpretative and reconstructive, based on empirical data of the context	Tools of current debates, motifs of actors, narratives, pieces of art, movies, …	Hermeneutic sciences and humanities such as sociology, linguistics (…)
Instrumental knowledge	Provision of a 'toolkit' for action and decision-making	Know-how about measures and instruments	Opportunities for intervention, governance of the system considered	Economic and political sciences, engineering sciences, law, medicine (…)

knowledge about possible sites for nuclear waste disposal on the one hand, and social science knowledge about participatory and legitimizing issues of social inclusion in the siting process. Both types of knowledge will occupy epistemic places in an overall assessment independent from each other.

However, things can be completely different. If an energy scenario were to be based on a macro-economic model using neoclassical presuppositions such as the *homo oeconomicus* model, whilst ethical considerations of energy consumers were considered in another part of the same assessment process based on an Aristotelian or Kantian model then there would be an inconsistency. Interference between incompatible models of human behaviour would occur, with implications difficult to assess.

The core issue is that usually no common frame of reference is available, consisting of a set of terms and their relations, theoretical elements and a well-known body of language. Without such a common frame of reference, however, proofs of consistency would simply not be possible. There might be hidden inconsistencies amongst the various notions, their definitions, the cognitive interests of the disciplines involved, their blind spots, presuppositions of the theories applied, and so forth. This is still an unsolved (and perhaps unsolvable) problem shadowing epistemic integration across borderlines of existing and often disciplinary frameworks of reference in TA and probably also beyond.

There can be no purely scientific solution to these challenges because (1) a common framework for classifying integrated knowledge as epistemologically sound, and in particular as consistent, is not available, and (2) because science has no mandate to make the final decision in cases of normative conflicts which cannot be integrated into a coherent picture. Rather, the process of integration must be regarded as a social and dialogical process, as an activity done by actors for a purpose and related to a specific context. Integration as part of the assessment process is bound to the requirements concerning TA's process quality ([1], Sect. 4.4.2). The only way to synthesize epistemologically incommensurable pieces of knowledge is to shape integration as dialogic communication and to control this dialogue with respect to the requirements mentioned, in particular with respect to transparency.

Often, scientific policy advice and decision-support close down spaces of decision-making to one solution, which is regarded as the best and presumably optimal one [39, 40]. For example, model-based approaches apply optimization algorithms and recommend the results to policymakers as the presumed objectively best one, assuming that 'science knows best' for shaping societal futures, e.g. in determining what should be done in areas such as the future of energy supply. However, according to its cognitive interest TA does not follow this approach. The dimension of inclusion renders it impossible to identify a 'one best solution' based on presumed scientific objectivity represented, e.g. by models and algorithms. The main reason is that the selection amongst values, political positions, images of humans and society, understanding of justice etc. [41] is beyond the mandate and expertise of science. It would be simply ideology to hide this observation behind a façade of presumed objectivity or of the epistemic authority of the science system. The obligation of TA of inclusion demarcates a deep difference to scientific optimization. Whilst *thinking in alternatives* is the appropriate mode of operation of TA, the selection amongst the alternatives is regarded as on behalf of the democratic decision-making procedures [16]. This position of TA demonstrates a deep difference concerning the understanding of integration in science. Whilst integration there, e.g. in integrative modelling, often aims at coherent optimization, integration in TA ends up in providing coherent alternatives of approaching the future.

6 Concluding Remarks

- Technology assessment has been enriching reflections on the technological advance and decision-making for more than fifty years, e.g. on issues of digitalization, energy technologies, sustainable development and biotechnologies
- TA has developed application fields in scientific policy advice, e.g. for parliaments, engagement in participation and public debate as well as becoming part of engineering and technology design.
- TA aims at enhancing reflexivity with respect to possible consequences of technology and by including various social and ethical perspectives.
- Integration of prospective knowledge provided by various disciplines but also the integration of knowledge, values, and interests of stakeholders and citizens belongs inherently to TA.
- Complexity management in TA needs integrated systems views for allowing reflected assessments of relevance of knowledge as well as of actors to be included.
- In contrast to integrating knowledge for optimization, e.g. in integrative modelling, integration in TA ends up in providing coherent alternatives of approaching the future as a basis for democratic deliberation.

References

1. Grunwald A (2019) Technology assessment in practice and theory. Routledge, Abingdon
2. Beck U (1992) Risk society. Towards a new modernity. Sage Publications, New Delhi. Original version: Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt a.M., Suhrkamp 1986
3. Beck U, Giddens A, Lash S (1994) Reflexive modernization. politics, tradition and aesthetics in the modern social order. Stanford University Press, Stanford
4. Delvenne P, Fallon C, Brunet S (2011) Parliamentary technology assessment as indications of reflexive modernization. Technol Soc 33(1–2):36–43
5. Rip A, Misa T, Schot J (eds) (1995) Managing technology in society: the approach of constructive technology assessment. Pinter Publishers, London
6. Bechmann G, Decker M, Fiedeler U, Krings B (2007) Technology assessment in a complex world. Int J Foresight Innov Policy 3:6–27
7. Owen R, Bessant J, Heintz M (eds) (2013) Responsible innovation: managing the responsible emergence of science and innovation in society. Wiley, London
8. Grunwald A (2014) The hermeneutic side of responsible research and innovation. J Responsible Innov 1(3):274–291
9. von Schomberg R, Hankins J (eds) (2019) International handbook on responsible innovation. A global resource. Edward Elgar, Cheltenham. https://doi.org/10.4337/9781784718862.00031
10. Meadows D, Meadows DH, Randers J (1972) The limits to growth: a report for the Club of Rome's project on the predicament of mankind. Universe Books, New York
11. Davies J (2016) The birth of the anthropocene. University of California Press, Oakland CA
12. Bijker WE, Law J (eds) (1994) Shaping technology/building society. MIT Press, Cambridge MA

13. Grunwald A (2012) Sustainability assessment of technologies—an integrative approach. In: Ghenai C (ed) Sustainable development—energy, engineering and technologies—manufacturing and environment, pp 35–62. InTech, Rijeka. https://doi.org/10.5772/26623
14. Habermas J (1970) Toward a rational society. Heinemann, London UK. Original version: Technik und Wissenschaft als Ideologie. Frankfurt a. M., Suhrkamp (1968)
15. von Schomberg R (ed) (1999) Democratizing technology. Theory and practice of a deliberative technology policy. International Centre for Human and Public Affairs, Hengelo
16. Grunwald A (2019) The inherently democratic nature of technology assessment. Sci Pub Policy 46(5):702–709. https://doi.org/10.1093/scipol/scz023
17. Abels G, Bora A (2016) Ethics and public participation in technology assessment. https://doi.org/10.13140/RG.2.2.35586.89282
18. Coenen C, Simakova E (2013) STS policy interactions, technology assessment and the governance of technovisionary sciences. Sci Technol Innov Stud 9(2):3–20
19. Rip A (2002) Co-evolution of science, technology and society. Expert review for the Bundesministerium Bildung und Forschung's Förderinitiative 'Politik, Wissenschaft und Gesellschaft' Berlin, Science Policy Studies
20. Bimber BA (1996) The politics of expertise in Congress: the rise and fall of the office of technology assessment. State University of New York Press, New York
21. Grunwald A (2006) Scientific independence as a constitutive part of parliamentary technology assessment. Sci Pub Pol 33(2):103–113
22. Cruz-Castro L, Sanz-Menendez L (2004) Politics and institutions: European parliamentary technology assessment. Technol Forecast Soc 27:79–96
23. Joss S, Bellucci S (eds) (2002) Participatory technology assessment—European perspectives. Westminster University Press, London
24. Barben D, Fisher E, Selin C, Guston D (2008) Anticipatory governance of nanotechnology: foresight, engagement, and integration. In: Hackett EJ, Amsterdamska O (eds) The handbook of science and technology studies. MIT Press, Cambridge, MA, pp 979–1000
25. Guston D (2014) Understanding 'anticipatory governance.' Soc Stud Sci 44(2):218–242
26. Bechtold U, Fuchs D, Gudowsky N (2017) Imagining socio-technical futures – challenges and opportunities for technology assessment. J Responsib Innov 4(2):85–99
27. Renn O (1999) Participative technology assessment: meeting the challenges of uncertainty and ambivalence. Futur Res Q 15:81–97
28. Wynne B (1995) Technology assessment and reflexive social learning: observations from the risk field. In: Rip A, Misa T, Schot J (eds) Managing technology in society. Wiley, London, pp 19–36
29. Jonas H (1984) The imperative of responsibility: in search of an ethics for the technological age. Chicago, University of Chicago Press. Original version: Das Prinzip Verantwortung. Versuch einer Ethik für die wissenschaftlich-technische Zivilisation. Frankfurt a. M., Suhrkamp
30. Adams B, Groves C (2007) Future matters: action, knowledge, ethics. Leiden, Brill
31. Selin C (2014) On not forgetting futures. J Responsib Innov 1(1):103–108
32. Konrad K, van Lente H, Groves C, Selin C (2017) Performing and governing the future in science and technology. In: Miller CA, Felt U, Fouché R, Smith-Doerr L (eds) The handbook of science and technology studies. MIT Press, Cambridge MA, pp 465–493
33. Grunwald A (2019) The objects of technology assessment. Hermeneutic extension of consequentialist reasoning. J Responsib Innov 7(1). https://doi.org/10.1080/23299460.2019.1647086
34. Grunwald A (2013) Modes of orientation provided by futures studies: making sense of diversity and divergence. Eur J Futures Stud 15:30–49. https://doi.org/10.1007/s40309-013-0030-5
35. Grunwald A (2012) Responsible nano(bio)technology: ethics and philosophy. Pan Stanford Publishing, Singapore

36. Hennen L (2005) Get the public involved? Participatory methods in technology assessment. In: Rohracher H (ed) User involvement in innovation processes. Strategies and limitations from a socio-technical perspective. München, Wien, Profil, pp 291–315
37. Rip A (1986) Controversies as informal technology assessment. Knowl Creation Diffus Utilization 8(2):349–371
38. Habermas J (1981) Theory of communicative action, two volumes. Beacon Press, Boston
39. Collingridge D, Reeve C (1986) Science speaks to power: the role of experts in policymaking. Frances Pinter, London
40. Stirling A (2008) Opening up and closing down: power, participation, and pluralism in the social appraisal of technology. Sci Technol Human Values 33(2):262–294
41. Wynne B (2016) Ghosts of the machine: publics, meanings and social science in a time of expert dogma and denial. In: Chilvers J, Kearnes M (eds) Remaking participation: science, environment and emergent publics. Routledge, London & New York, pp 99–120

Armin Grunwald is a Full professor of Philosophy at Karlsruhe Institute of Technology (KIT), Germany. He is a director of the Institute for Technology Assessment and Systems Analysis at KIT (ITAS) and Director of the Office of Technology Assessment at the German *Bundestag*.

His **professional backgrounds** include technology assessment, ethics of philosophy, theory of sustainable development, and the epistemology of inter- and transdisciplinary research.

In **his professional work**, he is member of several advisory commissions and committees in various fields of the technological advance.

In **his research work**, he is developing a theory of technology assessment in conjunction with applications in the fields of the transformation of the energy system, the ongoing digitalization, and new and emerging technologies such as synthetic biology and human enhancement.

He is an author and editor of multiple professional and research publications, amongst them 'Technology Assessment in Practice and Theory' (Routledge 2019) and 'The Hermeneutic Side of Responsible Research and Innovation' (Wiley 2016).

More about name and surname can be found Here (www.itas.kit.edu/english/staff_grunwald_armin.php).

The Artistic and Scientific Nature of Protein Structure: A Historical Overview

28

Julie B. Ealy

Sometimes progress is slow. But then there does come a time when a lot of people accept a new idea and see ways in which it can be exploited. And because of the larger number of workers in the field, progress becomes rapid. That is what happened with the study of protein structure.

Linus Pauling [1]

Summary

This author has been fascinated with proteins' visual beauty since her introduction to molecular modeling software for small molecules and subsequent use of computational software that permitted 3D visualization of proteins. With the use of computational software to view proteins for the first time, the visual possibilities resulted in a moment of awe with the seeming simplicity of a protein's primary structure of amino acids. These amino acids define a secondary structure wherein alpha helices and beta sheets are nested, fold and connect to form the tertiary structure, and form a quaternary structure such as ribonucleic acid (RNA) polymerase II, a complex of proteins. We can stand back, look at a protein as a piece of art, and know that its structure informs the viewer how the protein will function scientifically. The author has chosen several proteins of historical significance. Are there others? Yes, and so the initial question became, what proteins to choose. Enjoy the journey.

J. B. Ealy (✉)
Penn State Lehigh Valley, Center Valley, PA, USA
e-mail: jbe10@psu.edu

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Nesquehoning, PA, USA

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_28

Julian Voss-Andreae Heart of Steel (Hemoglobin) (Adapted from Wikimedia Commons, the free media repository https://upload.wikimedia.org/wikipedia/commons/6/68/Heart_of_Steel_%28Hemoglobin%29.jpg).

The code of this chapter is 01110000 01110010 01101111 01110100 01100101 01101001 01101110.

1 Introduction

Linus Pauling's was correct and rapid change is most likely an understatement. The hope is that this chapter will help the reader become more aware of how history and science have contributed to our knowledge of proteins. Thank goodness for those who seek to understand protein structure because when a crystal structure is solved, it provides information that can be helpful for other workers in the field. Protein structure contributes to our knowledge of how function, genetics, and medicine can support each other. The community of scientists who work together has and will continue to make rapid progress in the field.

There are four proteins discussed in this chapter: hemoglobin, DNA polymerase, RNA polymerase II, and severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) main protease. From the first time this author saw a cartoon representation of hemoglobin, there was an immediate need to know more about its structure and how that affected its function. How could one molecule play such a significant role in the delivery of oxygen to our tissues? DNA polymerase plays a critical role that is to pass on our genetic information to the next generation. Where would we be without it? RNA polymerase II receives a message from the genetic information present in DNA. Then transcription occurs along with messenger RNA formation, which leaves the nucleus to mediate protein formation through translation. The last protein chosen is SARS-CoV-2 main protease because of the pandemic in 2020 that killed more than 5.7 million people worldwide as of writing this chapter. The possibility exists that the protein structure could provide valuable assistance in developing a drug to reduce or even stop the pandemic.

2 Hemoglobin

Hemoglobin was chosen as one of the proteins discussed in this chapter because of its medical/scientific importance and artistic symmetry. It is responsible for the transport of oxygen and carbon dioxide within the body, and therefore it is essential to our survival. Its history began as early as 1840 with the discovery of crystalline material from earthworm blood and plate-like crystals of swine or human blood samples. This discovery was first described in a book published by Friedrich Ludwig Hunefeld [2]. It was not until 1864 that Felix Hoppe-Seyler named the crystals hemoglobin [3], and in 1870 Claude Bernard verified the role of hemoglobin as an oxygen carrier [4].

The scientific history of hemoglobin is not complete without Max Perutz's contribution, who won the Nobel Prize in chemistry in 1960 for the X-ray analysis of hemoglobin [5]. Though Max Perutz's intention for his Ph.D. thesis was to work with Frederick Gowland Hopkins on vitamins, his physical chemist professor, Herman Mark, arranged for him to meet crystallographer John Bernal. Even though Perutz protested because he knew nothing about crystallography, he drew inspiration from Bernal due to his vision that X-ray diffraction could solve large molecular structures even when sugar's structure had not been solved [6]. In 1938 Dr. G. S. Adair made crystals of horse hemoglobin and gave them to Max Perutz, John D. Bernal, and Isidor Fankuchen. As noted by the authors [7], the crystals were well suited for X-ray analysis. Deductive reasoning based on the research of other scientists [8, 9] and assuming there were two molecules per cell determined that the approximate molecular weight of hemoglobin without water was 69,000 g/mole. Chemical methods had shown a weight of 67,000 [7]. The actual molecular weight of hemoglobin as a tetramer is 64,458 g/mol. What an accomplishment this was! The molecular weight obtained by X-ray crystallography in 1938 was close to the actual mass.

In 1960, Perutz became the winner of the Nobel Prize for the 5.5 Å X-ray crystal structure of hemoglobin. He remarked that the structure provided a first glimpse of the structure and said, "…but it may well be necessary to solve the structure of at least one of the two forms in atomic detail. Due to the enormous amount of labor involved, this may take some time, but not much, perhaps, compared to the 22 *years* needed for the initial analysis" (NobelPrize.org. Nobel Media AB 2020. 653–673. https://www.nobelprize.org/prizes/chemistry/1962/perutz/lecture/).

Thankfully, Perutz did not settle for the 5.5 resolution of hemoglobin but obtained a 2.8 Å resolution in 1968 [10] and a 1.74 Å resolution in 1984 [11].

2.1 Structure

Figure 1 shows the artistic symmetry of the quaternary structure of hemoglobin with the alpha helices of the alpha and beta chains appearing to be intertwined like the old roots of a tree wrapped around the base without knowing where one part

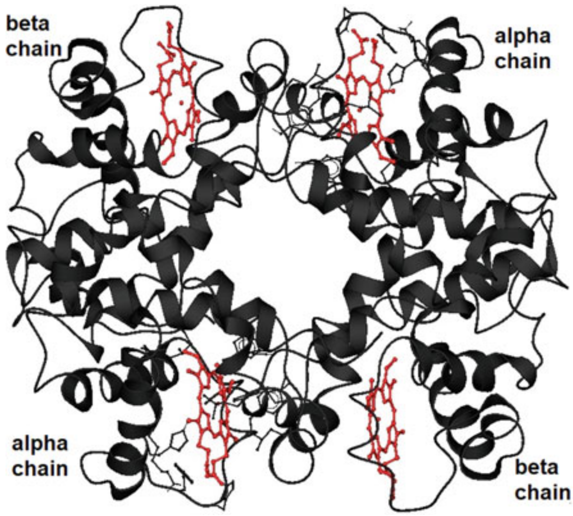


Fig. 1 A view from the top of the quaternary structure of hemoglobin with the alpha and beta chains in black and the heme group in red [11]. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from John Wiley and Sons)

begins, and another part ends; within the structure, the four heme groups appear suspended vertically in space. The heme group is represented progressively from Figs. 2, 3 and 4. Figure 2 shows the heme group turned around, so it is face-to-face with the viewer with iron anchored in the center and intermolecular hydrogen bonds from nitrogen to iron. Still face-to-face but shown in the protein's active site, Fig. 3 displays not only the four hydrogen bonds from nitrogen to iron, but there is also an additional hydrogen bond from histidine 87 of the protein to iron. The five hydrogen bonds provide stability so the heme group can anchor itself in the protein's active site. Hemoglobin is responsible for transporting oxygen and carbon dioxide within the body. Figure 4 represents oxygen covalently bonded to the iron to complete an octahedral complex from the six ligands. A magnificent protein indeed!

2.2 When Serendipity Happens

There are various diseases associated with hemoglobin. Here, the main focus is sickle cell anemia. It was Vernon Ingram who discovered that a single genetic mutation caused sickle cell anemia. In 1952, Ingram found himself, after a challenging job search, in Max Perutz's laboratory where his discovery would take place. Perutz wanted a heavy atom placed in a specific location in his hemoglobin crystals and Ingram completed the task for Perutz. At that point, Ingram wanted to know if the heme group is linked with amino acid side chains. However, the preparation of tryptic hemoglobin peptides with the obvious heme group did not amount to much because there were no specific bonds from hemoglobin to the heme group. Serendipity knocked on the door when Tony Allison came to the laboratory and brought with him samples of sickle cell anemia hemoglobin. Max Perutz and

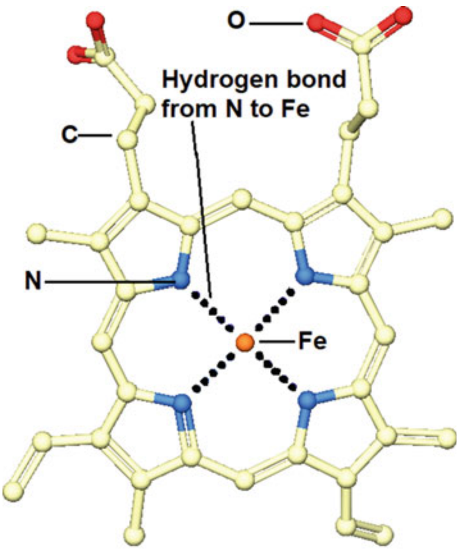


Fig. 2 The heme group protoporphyrin IX composed of four pyrrole subunits with four hydrogen bonds from N to Fe in the center [11]. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from John Wiley and Sons)

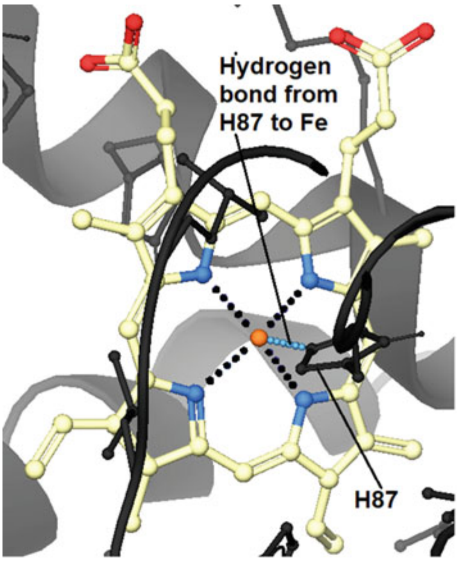


Fig. 3 The heme group in the active site of deoxygenated hemoglobin. In addition to the four hydrogen bonds from nitrogen to iron there is also a hydrogen bond from H87 to iron [11]. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from John Wiley and Sons)

Frances Crick suggested that normal and sickle cell hemoglobin should be compared. Ingram working with Allison argued that if there was a difference of only one peptide that was the only peptide in the protein that was different, and therefore the rest of the peptide sequence was normal. Their initial assumptions were later proven correct; that is, only one amino acid difference caused sickle cell anemia.

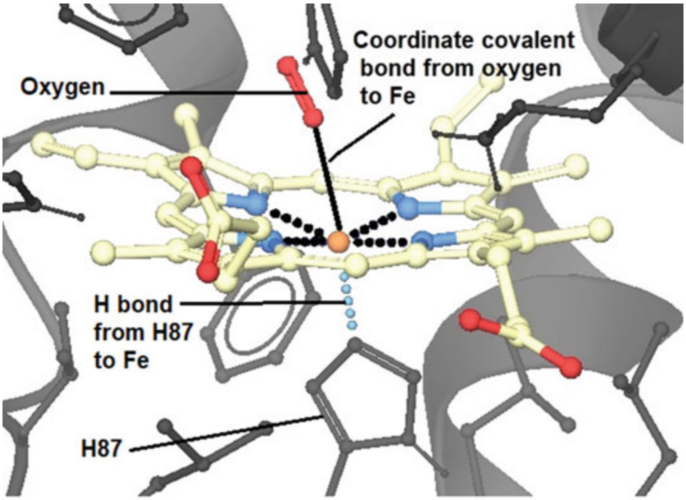


Fig. 4 A representation of oxygenated hemoglobin with a coordinate covalent bond from oxygen that can reversibly bond to Fe. An octahedral complex of six ligands is formed (4HHB.pdb.) [11]. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from John Wiley and Sons)

Ingram's continued research at MIT with many graduate students resulted in his conclusion "that we were chemically exploring the mechanisms of Mendelian inheritance and the evolution of gene clusters" [12, p. 3].

2.2.1 Sickle Cell Anemia

In a 1951 speech, Linus Pauling talked about his 1949 discovery that an abnormal form of hemoglobin prevalent in Negroes accounted for the expression of sickle cell anemia disease. He indicated that about 8 percent of American Negroes have sickle cell anemia with no issues, but approximately 3 percent suffer from severe anemia [13]. According to the centers for disease control and prevention (CDC), the present estimate for black or African American births is approximately 0.3% [14]. In sickle cell anemia, red blood cells are not oxygenated because they twist into a shape like a half-moon, lose their flexibility, become sticky, and have difficulty passing through the capillaries; the capillaries become crowded, resulting in a lack of oxygen. The spleen and leucocytes remove the sickled cells and the person becomes anemic. The red blood cells from sickle cell anemia patients were carefully studied. Their red blood cells contained abnormal hemoglobin, which led to the conclusion that sickle cell anemia was a molecular disease. This was the first clarification of disease as molecular [15]. Pauling was fascinated by sickle cell anemia and commented that "Hemoglobin is one of the most interesting chemical substances in the world – to me it is the most interesting of all" [15, p. 556].

2.3 Artistic Sculpture

The graphical analysis at the beginning of the chapter shows Julian Voss-Andreae's sculpture of hemoglobin. His sculptures are available in North America, Asia, Europe, Africa, and Australia. He completed graduate research in quantum physics and its mysteries. His expertise in diverse fields of science provides a continual source of inspiration for his sculptures.

The sculpture's metamorphosis is shown from the left after the unveiling, ten days later, and several months after exposure to the elements when it had acquired a deeper coloration as it became saturated with oxygen. The blood-red glass sphere in the center represents a drop of blood [17]. As we age over time, so did the sculpture. In the words of the artist: "But the act to pull the individual protein structure out of science and putting it into an art context and thus allowing the emotions to become part of the perception feels like a step in the right direction" (Personal communication with Julian Voss-Andreae).

3 DNA Polymerase

DNA polymerase is what makes it possible to pass our genetic information on to the next generation. The enzyme's responsibility is to duplicate the double-stranded DNA containing the genetic code. Arthur Kornberg became the winner of the Nobel Prize in physiology or medicine in 1959 for deoxyribonucleic acid biologic synthesis. He said, in his Nobel Prize lecture that DNA appears as a tape recorder that "carries a message in which there are specific instructions for a job to be done … and like a tape recording, exact copies can be made from it so that this information can be used again and again and elsewhere in time and space" [18, p. 665]. In his Nobel Prize presentation, he asked his audience to assume DNA is the genetic substance that can be copied and thereby provides for the development of many generations of offspring and directs protein synthesis [18]. DNA polymerase is a nearly perfect enzyme because it creates an exact copy of DNA with fewer than one mistake in a billion bases with cytosine matched to guanine and adenine matched to thymine (Fig. 5) [16]. Moreover, the enzyme proofreads after each base, moves backward by one base pair of DNA, and then excises the incorrect base pair [19]. Indeed, DNA polymerase has proposed this process as a mistake eraser.

3.1 Never in a Vacuum

Outstanding accomplishments do not happen in a vacuum, and the story of Arthur Kornberg receiving the Nobel Prize in 1959 is no exception. In 1955, the discovery of DNA polymerase occurred when purifying *Escherichia coli* enzyme. The

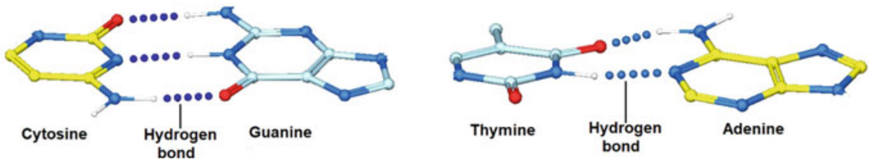


Fig. 5 The two pairs of DNA nucleotides, cytosine paired with guanine and thymine paired with adenine. Three hydrogen bonds connect cytosine to guanine and two connect thymine with adenine. The hydrogen bonds between the nucleotides provide stability to the DNA helix but the bonds also unzip during replication. The image was created with ICM-Pro (Molsoft L.L.C.) [16]. (*Source* Copyright from Springer Nature Customer Service Centre GmbH)

triphosphates of the four deoxynucleosides were mixed and underwent incubation with ATP and extracts of *E. coli.* Though a minute amount of enzyme was produced, Kornberg remarked that "we tried to drive a wedge, and the hammer was enzyme purification" [16, p. 670]. The co-authors on the early 1958 publications for the synthesis of DNA polymerase were Kornberg, I. R. Lehman, Ernie Simms, and Maurice Bessman [20]. Two manuscripts from the previously named authors were rejected by the *Journal of Biological Chemistry* (JBC) in 1957 that described the 1955 research, one aspect of which was that DNA polymerase could be partially purified. Critical comments from the Editor-in-Chief were: "It is very doubtful that the authors are entitled to speak of the enzymatic synthesis of DNA"; "Polymerase is a poor name"; "Perhaps as important as the elimination of certain banalities…" [20, p. 34736]. Fate stepped in, and a new Editor-in-Chief accepted the papers for 1958 publication in JBC. Another publication that polymerase copies the DNA template, was accepted in the *National Academy of Sciences Proceedings* in 1958 [20].

Arthur Kornberg wrote an article, "Remembering Our Teachers," and talks about two teachers who greatly influenced him [21]. As teachers, we hope our students will remember us, and just as Kornberg served as a mentor and teacher to those who were co-authors on the 1958 JBC articles, so was he mentored by Carl and Gerty Cori, who became his most devoted patrons. The discovery of glycogen phosphorylase by the Cori's inspired Kornberg to seek an enzyme that could be responsible for DNA synthesis, but not the DNA chain's base pairing described by Watson and Crick. Kornberg believed that the most effective way to understand biological events was through enzymology. Another significant teacher of Kornberg's was Severe Ochoa, with whom he shared the Nobel Prize in physiology or medicine in 1959 and with whom he collaborated for a year in 1947. It was there that he learned the basics of biochemistry, enzyme fractionation, and spectrophotometry. He remarks about his time in Ochoa's laboratory as the "most exciting in my life" [21, p. 9] and about enzymes, "…I have never met a dull or disappointing one" [21, p.9]. A "Holy Toledo" event, as described by Kornberg, occurred after the completion of a very large-scale preparation based on livers taken from several hundred pigeons. The final enzyme fraction fell over and all of it spilled onto the floor. Though Kornberg was devastated, Ochoa was reassuring. For some reason, Kornberg had saved a fraction believed to be inactive, but it contained the bulk of the enzyme activity when tested the following morning. The calling out of "Holy Toledo" by Kornberg was a source of amusement by Ochoa and one of relief and pleasure by Kornberg [21].

3.2 Replication of DNA

DNA polymerase's function is to synthesize DNA from deoxyribonucleotides. Before a new copy of DNA originates, an enzyme named helicase breaks the hydrogen bonds between the nucleotides, i.e., between guanine and cytosine and between adenine and thymine. It allows the helix to open and provides two single-strands that act as guides during the replication. DNA polymerase cannot start a new chain, and primase performs the synthesis of two RNA bases, which will serve as a primer to add two nucleotides onto a pre-existing 3′-OH group. Later at a 5′exonuclease site, the RNA fragments are removed. Each nucleotide base in a template strand is matched to the complementary base in a free nucleotide. Because nucleotides can only be added to the 3′ end of a new strand, the strand's elongation is only in the 5′–3′ direction of the newly formed daughter strand. Correction of mistakes, or proofreading, occurs by the action of the 3′–5′ exonuclease domain of DNA polymerase; this is to assure removal of an incorrect base pair. Topoisomerase serves a unique role when DNA overwinds or underwinds during replication because it can cut the phosphate backbone so the DNA strand can unwind; the backbone then reseals [22–25] (Fig. 6).

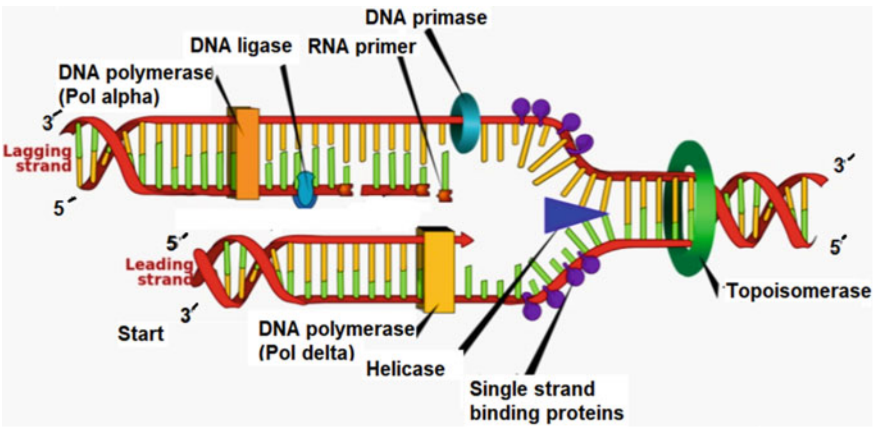


Fig. 6 The figure represents the various actors and enzymes involved in DNA replication [22]

3.3 Structure of DNA Polymerase

Many but not all DNA polymerases have a shape described as a hand with the space between the fingers and the thumb as the perfect location for the DNA helix (Fig. 7). What does reflect this shape is 1tau.pdb (Fig. 8) [16].

The DNA polymerase of 1tau.pdb was isolated from *Thermus aquaticus* in 1976 [26] (Fig. 8). A close-up of the DNA with the template and newly built DNA strand is shown in Fig. 9. One domain that 1tau.pdb does not have is the 3′–5′ exonuclease proofreading domain of DNA polymerase. It is not understood why but the accepted explanation is that it lives in hot springs such as Yellowstone National Park, and perhaps the heat performs the proofreading [19]. The DNA polymerase of *Thermus aquaticus* has been renamed *Taq* polymerase or *Taq* pol and is now used in the polymerase chain reaction (PCR) to amplify short DNA segments [27]. Because high temperatures that can denature proteins are required during PCR, *Taq* pol's ability to withstand high temperatures resulted in the replacement of *E. coli* DNA polymerase that had been used in PCR [28].

To end this section, a quote from H. Theorell's presentation speech at the Nobel Prize award ceremony for Kornberg and Ochoa provides a somewhat poetic description: "Their molecules are very large and are built up from thousands of smaller units linked together in chains—just like a string of pearls- which have a

Fig. 7 Representation of DNA polymerase as a hand with the DNA nestled in the space between the fingers and the thumb [16]. (*Source* Copyright from Springer Nature Customer Service Centre GmbH)

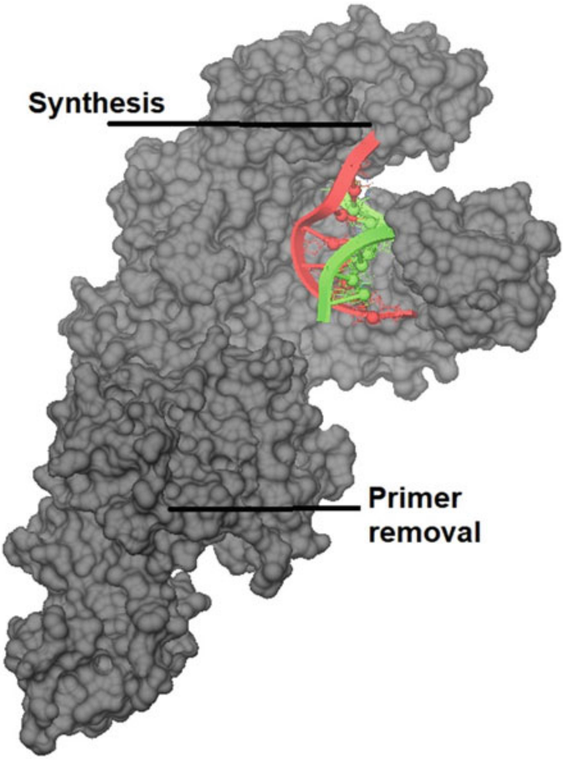


Fig. 8 A space-filling cartoon visual with DNA in the region between the "fingers" and thumb. The image was created with ICIM-Pro (Molsoft. L.L.C.) [16]. (*Source* Copyright from Springer Nature Customer Service Centre GmbH)

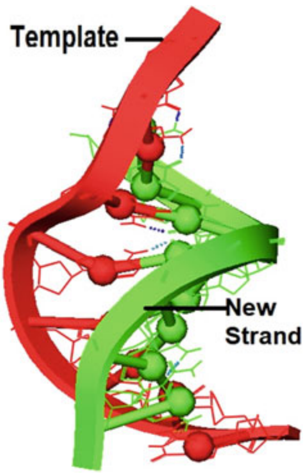


Fig. 9 Close-up DNA with the template strand in red and the new strand in green. The image was created with ICM-Pro (Molsoft L.L.C.) [16]. (*Source* Copyright from Springer Nature Customer Service Centre GmbH)

tendency to form helices. Single helices join together in complicated threads that can contain proteins or nucleic acids or both. In the mixed super molecules, the reactions of life proceed in the subtle pattern of the intimately associated strands" [29].

3.4 When Something Goes Wrong Genetically

As discussed earlier, sickle cell anemia results from a mutation in the hemoglobin subunit beta (HBB) gene, making the beta-globin strand of hemoglobin specifically. An abnormal beta-globin, the so-called hemoglobin S, replaces both beta-globin subunits in hemoglobin. Anemia can result from the early death of sickle-shaped red blood cells. Another mutation of the HBB gene occurs when there are low beta-globin levels, resulting in beta-thalassemia [30]. Thalassemia limits the amount of hemoglobin a person produces, and consequently, severe anemia can occur [31].

Down syndrome is not due to a gene mutation but results when chromosome 21 undergoes an extra copy. A person with Down syndrome may exhibit reduced cognitive function with a higher disposition for congenital heart defects. Cystic fibrosis is another genetic disorder with a higher probability of inheritance when both parents have the cystic fibrosis gene. Systems affected by this genetic disorder are respiratory, digestive, and reproductive [32]. A mutation on chromosome 7 is responsible for cystic fibrosis, and there must be a copy of the CF gene from each parent [32].

Last but not least is breast cancer that one in eight women will develop in their lifetime. Breast cancer is not restricted to women, with approximately 1 in 2000 men diagnosed with breast cancer. Though most breast cancer cases are not inherited, mutations in the BRCA1 or BRCA2 gene increase the chance that cancer will develop, and the mutated gene can be passed on by either the mother or father [33].

3.5 A Blend of Science, Computer Graphics, and Art

James Hamlin and Carlo H. Sequin worked together and used computer graphics to design ribbed sculptures such as the double helices shown in Fig. 10. There were two guide rails used that twisted around one another several times to form a parabolic arch designed to take off upward. The images were modeled with computer graphics with the result, an artistic sculpture of the DNA helix. The authors experimented with two abstraction levels that led to a "fertile domain of intriguing geometrical shapes, which may make informative mathematical models as well as exciting monumental tubular sculptures" [34].

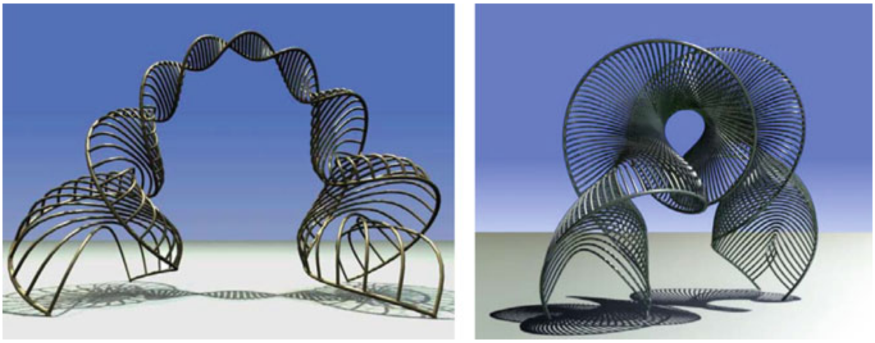


Fig. 10 "Computer-Generated Ribbed Sculpture" by James Hamlin and Carlo H. Séquin, CS Division, University of California, Berkeley, 2010. (*Source* Copyright from [34])

4 RNA Polymerase II

What is a gene? A single gene is only a portion of a chromosome that contains a single DNA molecule that is passed from parents to their offspring and could be a few hundred or thousands of base pairs. Genes code for a protein that has a specific function, but one gene can also produce multiple proteins, and some genes do not code for any protein but code for RNAs [35]. What is the role of RNA polymerase II in this process and the formation of a protein? Its function is to transcribe messenger RNA (mRNA) by adding ribonucleotides to a growing RNA strand that corresponds to complementary deoxyribonucleotides on a single-strand of DNA, called the template strand. Similar to DNA base pairing, cytosine is the complement of guanine in RNA and vice versa. However, though thymine of DNA pairs with adenine of RNA, adenine of DNA requires uracil of RNA [36]. The mRNA that is transcribed from RNA polymerase II leaves the nucleus and in the cytoplasm relays the message for the creation/translation of a specific protein [37, 38].

4.1 Structure and Process of Transcription

RNA polymerase II (Pol II) (Fig. 11) comprises ten domains plus two more domains to form a 12-domain complex [39], though not all 12 are shown in Fig. 12 [40]. Perhaps the complexity and beauty of Pol II are related to its role in the formation of messenger RNA (mRNA) that carries the message for the construction of proteins essential to life, thereby can serve as a drug target for disease and can be affected by a genetic mutation.

The transcription initiation that will result in mRNA formation relies on an architectural masterpiece complex. Pol II acts as a mediator and contributes to this complex along with transcription factors such as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH that account for transcription events that critically involve Pol II.

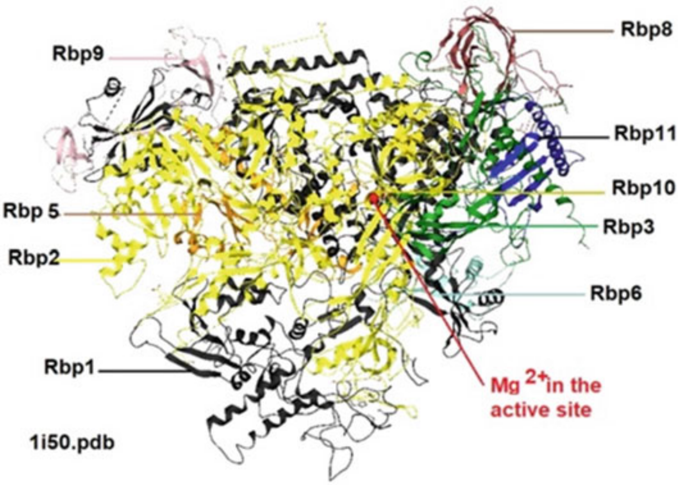


Fig. 11 RNA polymerase II at 2.8 angstroms [40]. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from The American Association for the Advancement of Science)

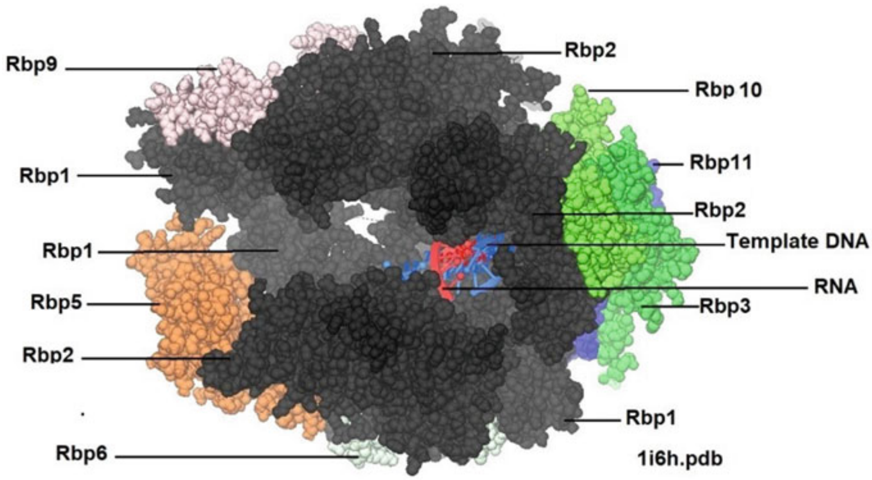


Fig. 12 An RNA polymerase II elongation complex at 3.3 angstroms, with the DNA-RNA hybrid deep in the active site of the complex [40]. The template DNA strand is shown in blue and the RNA strand in red. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from The American Association for the Advancement of Science)

Together these factors form the pre-initiation complex (PIC). The mediator complex serves to control the PIC activity while it interacts with Pol II [40]. The transcription process begins when a closed promoter complex containing the 10-domain Pol II core, a Pol II subcomplex Rbp 4/7, and various transcription factors form.

The role of Rbp 4/7 is to stabilize the Pol II clamp so it maintains a closed conformation that will only permit single-stranded DNA to pass through to the active site. The complete Pol II-TFIIF complex begins to form when Pol II binds to TFIIF [38].

The Pol II-TFIIF complex with the TBP-TFIIB-DNA complex initiates a closed promoter complex. There, TFIIB builds a bridge between Pol II and the promoter, and docking between the two complexes results from the interaction of the TFIIB N-terminal domain with the Pol II dock domain and the C-terminal TFIIB core domain with the wall of Pol II. The N- and C-terminal domains of TFIIB are connected by the TFIIB reader and linker regions extending through the Pol II cleft. When the DNA strands unwind and separate, the closed promoter complex opens, and the template single-strand DNA positions itself in Pol II's active site where RNA synthesis can be initiated. Initiation factors are released when RNA reaches a certain length with subsequent formation of a stable elongation complex (EC). Then, the EC, a DNA-RNA hybrid possessing eight to nine base pairs, continues to interact with a nucleotide of the growing messenger RNA strand [40, 41].

5 Medicine, Proteins, and Genes: Pharmacogenetics

With the sequencing of the human genome in 2003, researchers began to question whether genetic information could help choose the treatment for different genetic disorders. In the 1950s, pharmacogenetics developed because individuals with different genetic makeups responded differently to drugs. One application of pharmacogenetics uses a diagnostic test to check for different levels of a particular gene in breast cancer. Some breast cancer patients produce HER2 proteins, with HER2 standing for human epidermal growth factor 2 located on chromosome 17 [42]. A miracle drug, Herceptin, has been developed for women who produce HER2 proteins. Unfortunately for those whose tumor genes do not produce the HER2 proteins, the drug does not work [43].

Administration of a neuromuscular blocker such as succinylcholine can result in it taking three to eight times longer to regain neuromuscular function than the normal 5–10 min [44]. This results from variations in the gene for a particular enzyme, plasma butyrylcholinesterase, located on chromosome 3 [45].

Some people who have colon cancer take the drug Camptosar. Suppose they have a variant form of the proteins that metabolize the drug. In that case, they can develop diarrhea and a life-threatening infection, which suggests that pretesting for the gene that affects the proteins would be beneficial. Those proteins are uridine diphosphate glucuronosyltransferases (UCT) with the UGT1A gene mapping to chromosome 2 [46]. When proteins function normally, they render certain compounds water-soluble with biliary and renal elimination [46].

Warfarin acts as an anticoagulant indicated to block the formation of blood clots. However, the same dose does not have the same effect on people. Warfarin targets a particular protein in the blood and if there is genetic variation in genes VKORC1 and CYP2C9 located on chromosome 10, the activity of the targeted protein can be affected differently [47]. Therefore, it would be beneficial to design a genetic test to provide each person's specific dose based on their genetic makeup [43].

6 SARS-CoV-2 Main Protease

The choice of this protein and its connection with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) are significant, and the pandemic caused by a novel coronavirus cannot be overlooked in history. On February 11, 2020, SARS-CoV-2 was named coronavirus disease (COVID-19) [48]. Shortly after, on March 11, 2020 [49], the World Health Organization (WHO) issued a global pandemic due to COVID-19. COVID-19 seems to share 88% identity with SARS-like coronaviruses that occur in bats. This implicates bats as the reservoir hosts, though there may have been an intermediate host. COVID-19 was divergent enough from SARS-CoV that emerged in November 2002 to be considered a new coronavirus [50].

Previous research on respiratory infections due to coronaviruses SARS-CoV and the Middle East respiratory syndrome (MERS-CoV) that respectively happened to humanity in 2002 and 2012 might help develop a vaccine or antiviral drugs against COVID-19 [49]. Coronaviruses are positive-sense, single-stranded RNA viruses that can serve as messenger ribonucleic acid (mRNA) and can be translated into polyproteins such as SARS-CoV-2 main protease, M^{pro} [51] that is essential in the coronaviral life cycle. Like putting a glove over your hand, M^{pro} (6lze.pdb) superimposed over bat protease HKU4-CoV 3CL(pro) (4yoj.pdb) illustrates an excellent fit between the two proteins (Fig. 13) [51, 52]. Their striking structural similarity demonstrates why bat-derived HKU4 3C-like protease was used to suggest MERS-CoV inhibitors [52] and further verify bats as reservoir hosts of COVID-19. Knowledge of the 3D structure of M^{Pro} combined with computational studies [51, 53], experimental data [50], and lessons learned from historical case studies of different proteins [54] have the potential to provide helpful information for drug candidates for COVID-19. To address the last protein's artistic and scientific nature in this chapter, M^{pro}, the focus will be on several computational software visuals. Together the representations will hopefully show that a picture is worth a thousand words.

6.1 Superimposition of M^{pro} on HKU4-CoV 3CL(Pro)

How do you know where to begin when a drug or vaccine is needed for a disease such as COVID-19? Is there a specific protein associated with the disease that might be a good candidate to target for drug development? Because three-dimensional knowledge of protein structure provides useful information about its function, a good place to begin with for COVID-19 was with bat coronavirus because COVID-19 has an 88% identity shared with SARS-like coronaviruses occurring in bats [50]. More specifically, because HKU4 3C-like protease was used to suggest MERS-CoV inhibitors [52], was it also structurally similar to SARS-CoV-2 main protease? Like placing the palm of your hands together, it is evident they are structurally similar when the three-dimensional image of M^{pro} (6lze.pdb) is superimposed on HKU4-CoV 3CL(pro) (4yoj.pdb), Fig. 13.

6.2 Aldehyde 11a in the Active Site of M^{pro} (6lze.pdb)

Two compounds designated 11a and 11b that could potentially inhibit the active site of M^{pro} (6lze.pdb) were designed and synthesized [51]. A simple representation of binding and how a potential drug fits into an active site is shown in Fig. 14. It is similar to two pieces of a puzzle that work together. However, with respect to a potential drug and a protein, they do not always fit perfectly.

Aldehyde 11a is presented as one example of a molecule that warranted further study as a potential drug against M^{pro} (6lze.pdb) [51]. Its placement (magenta) is shown in the active site of M^{pro} (6lze.pdb) (black) (Fig. 15) with the surrounding active site amino acids of the protein shown in yellow. Figure 16 shows a close-up of aldehyde 11a where it is nestled in the active site of M^{pro} (6lze.pdb) like a baby in a cradle (Fig. 16). Not shown are various intermolecular hydrogen bonds,

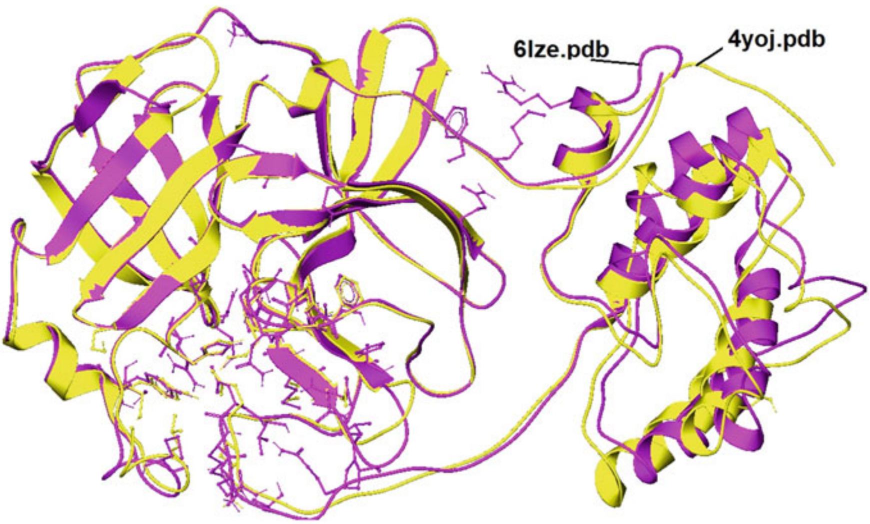


Fig. 13 Superimposition of SARS-CoV-2 main protease [51] (magenta) and HKU4 3C-like protease [52] (yellow). The image was created with ICM-Pro (Molsoft. L.L.C.). (*Source* Copyright from The American Association for the Advancement of Science and Elsevier)

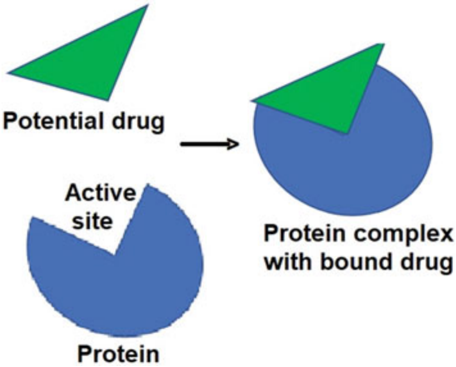


Fig. 14 Simplified representation of a potential drug and its ultimate location in the active site of a protein

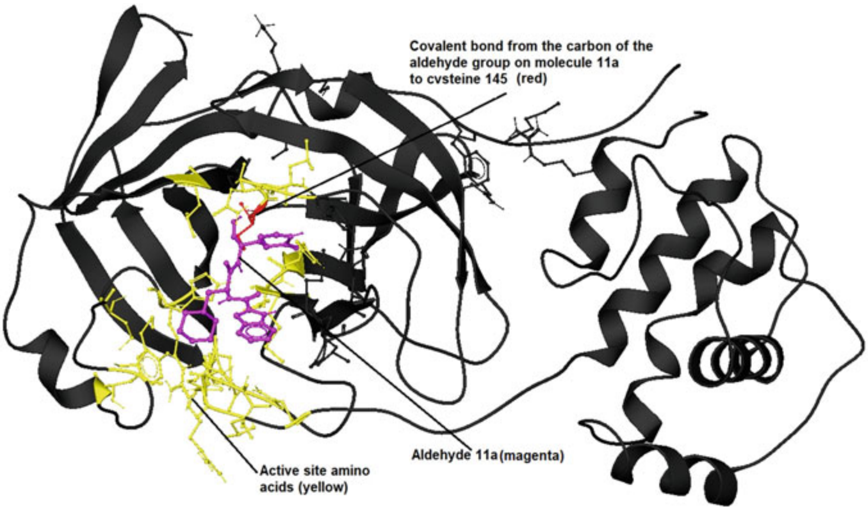


Fig. 15 Structure of SARS-CoV-2 main protease, M^{pro}, (black) complexed with an aldehyde structure, 11a, (magenta) that is located in the active site (yellow) of the protein [51]. A covalent bond to cysteine 145 is shown in red. The image was created with ICM-Pro (Molsoft L.L.C.). (*Source* Copyright from The American Association for the Advancement of Science)

hydrophobic interactions, but a standard covalent bond (shown in red, Fig. 15) from the carbon of the aldehyde group on molecule 11a to C145 is shown. All of these contribute to the stabilization of the molecule in the active site [51].

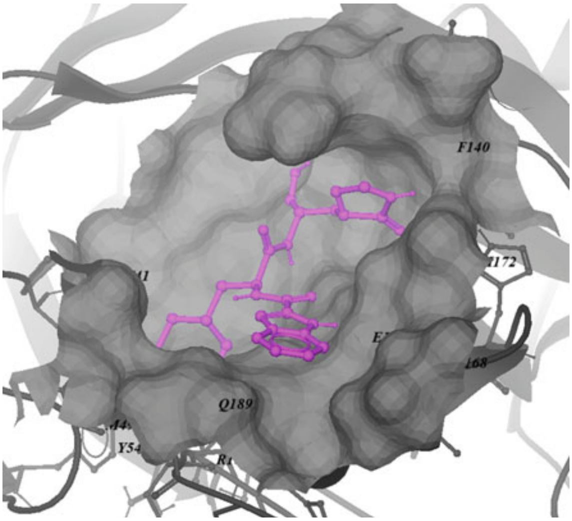


Fig. 16 Aldehyde 11a nestled in the active site of M^{pro} [51]. The image was created with ICM-Pro (Molsoft. L.L.C.). (*Source* Copyright from The American Association for the Advancement of Science)

6.3 Wet Bench Data—Aldehyde 11a

In addition to in silico data to support aldehyde 11 as a candidate for further study, the IC_{50} value showed excellent inhibitory strength. The molecule was also evaluated in vitro and exhibited a good antiviral effect on M^{Pro}. Injection of the aldehyde intraperitoneally and intravenously in mice demonstrated good bioavailability. In vivo toxicity results in rats and beagle dogs with once-daily dosing by intravenous drip resulted in no obvious toxicity [51].

6.4 Complementarity

The specific molecule is chosen to demonstrate the possibility of a molecule that may be a good candidate for further clinical studies that combine computational and wet bench scientific data. The in silico computational information demonstrates that the three-dimensional nature of a protein can speak to us like a fine piece of art and subsequently suggest ways that can inform drug design. The wet bench/experimental data often provides the only support for a potential drug, but they complement each other scientifically.

7 Conclusion

This chapter has been an adventure into proteins' artistic and scientific nature with many segues into the historical literature. Historically the advent and sophistication of computational software made the analysis of protein crystallographic data visually accessible and useful to scientists. Even non-scientists occasionally see cartoon structures or models of proteins in the news [55–57]. The medical profession has benefitted tremendously by understanding and diagnosing genetic disorders related to proteins and disease. When the human genome was solved, pharmacogenetics's birth made us more aware that *not all people react in a similar fashion to the same* dosage of medicine; it became important to ask whether genetics was responsible for the differences. Amazing connections among many scientists who contributed to each of the proteins in this chapter were a surprise and a delight to discover and emphasize that communication and sharing ideas do not occur in a vacuum. And then, there are the artists who combine their artistic talent and scientific knowledge to present us with a 3-D sculpture that illustrates the beauty of science. What more could we ask for?

Core Messages

- Scientific history can and should continue to influence future scientific decisions.
- Multi-disciplinary scientists need to communicate and cooperate to ensure future progress.
- The structure of a protein informs us how it functions, and this information helps make personalized drug decisions.
- Computational software has provided protein structure information that promotes a better understanding of diseases and drug design advances.
- A genetic mutation might underlie disease susceptibility and also an individual's reaction to drugs.

References

1. Pauling L (1986) Crossing the boundaries of science. From an interview with Neil A. Campbell. Bioscience 36(11):737. https://doi.org/10.2307/1310282. https://www-jstor-org.ezaccess.libraries.psu.edu/stable/1310282
2. Hunefeld FL (1840) Der Chemismus in der thierischen Organisation. Physiologisch-chemische Untersuchungen der materiellen Veraanderungen, oder des Bildungslebens im thierischen Organismus; insbesondere des Blutbildungsprocesses, der Natur der Blut Korperchen und ihrer Kernchen. Ein Beitrag zur Physiologie und Heilmittellehre.. Brockhaus, Leipzig (The chemistry in the animal organization: physiologically—chemical tests of material changes or educational life in the animal organism, especially the formation of blood Processes, the nature of the blood cells and their Kernchen : a contribution to physiology and medicine teaching)
3. Hoppe-Seyler F (1864) Ueber die chemischen und optischen Eigenschaften des Blutfarbstoffs, Archiv f¨ur pathologische Anatomie und Physiologie und f¨ur klinische Medicin Virchows Arch. 29(5–6):233–235 (About the chemical and optical properties of the blood pigment … Archive for pathological anatomy and physiology and for clinical medicine)

4. Bernard C (1857) Lecons sur les Effets des Substances Toxiques et M´edicamenteuses. Bailliere, Paris, France. https://www.biodiversitylibrary.org/bibliography/1834#/summary (Lessons on the effects of toxic and drug substances)
5. Perutz MF (1962) Nobel Lecture. NobelPrize.org. Nobel Media AB 2020:653–673. https://www.nobelprize.org/prizes/chemistry/1962/perutz/lecture/. Retrieved July 16, 2020 from https://www.nobelprize.org/uploads/2018/06/perutz-lecture.pdf
6. Eisenberg D (1994) Max Perutz's achievements: How did he do it? Protein Sci 3:1625–1628. https://doi.org/10.1002/pro.5560031001
7. Bernal J, Fankuchen I, Perutz M (1938) An X-ray study of chymotrypsin and hæmoglobin. Nature 141:523–524. https://doi.org/10.1038/141523a0
8. Chick H, Martin CJ (1913) The density and solution volume of some proteins. Biochem J 7:92–96. https://doi.org/10.1042/bj0070092
9. Haurowitz F (1924) Zur Chemie des Blutfarbstoffes. 1. Mitteilung. Über einige krystallisierte Hämoglobinderivate und über das sogenannte Krystallwasser derselben. Z. Physiol. Chem.136:147–159. The chemistry of the blood pigment. 1st communication. About some crystallized hemoglobin derivatives and their so-called crystal water. https://doi.org/10.1515/bchm2.1924.136.3-4.147
10. Perutz MF, Muirhead H, Cox JM, Goaman LCG (1968) Three-dimensional fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. Nature 219:131–139. https://doi.org/10.1038/219131a011
11. Fermi G, Perutz MF, Shaana B, Fourme R (1984) The crystal structure of human deoxyhaemoglobin at1.74Å resolution. J Mol Biol 175(2):159–174. https://doi.org/10.1016/0022-2836(84)90472-8. Image of 4HHB created with ICM-Pro, Molsoft L.L.C.
12. Ingram VM (2004) Sickle-cell anemia hemoglobin: the molecular biology of the first "Molecular Disease"—the crucial importance of seredipity. Genetics 167:1–7. https://doi.org/10.1534/genetics.167.1.1
13. Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. Science 110(2865):543–548. https://doi.org/10.1126/science.110.2865.543
14. CDC (2019) https://www.cdc.gov/ncbddd/sicklecell/data.html. Retrieved 19 July 2020
15. Pauling L (1952) The hemoglobin molecule in health and disease. Proc Am Philos Soc 96:556–565. Downloaded from http://www.jstor.com/stable/3143656. Retrieved 20 June 2020
16. Eom SH, Wang J, Steitz TA (1996) Structure of Taq polymerase with DNA at the polymerase active site. Nature, 382:278–281. Image of 1TAU created with ICM-Pro, Molsoft LLC. Figures 6, 9–10. https://doi.org/10.1038/382278a0
17. Voss-Adreae J (2013) Unraveling life's building blocks: sculpture inspired by proteins. Leonardo 46 (1):12–17. www.jstor.org/stable/23468111. Retrieved 18 July 2020
18. Kornberg A (1959) The biologic synthesis of deoxyribonucleic acid. Nobel Lecture, December 11, 1959. https://www.nobelprize.org/uploads/2018/06/kornberg-lecture.pdf. Retrieved 20 July 2020
19. PDB DNA polymerase molecule of the Month. https://pdb101.rcsb.org/motm/3. Retrieved 19 July 2020
20. Lehman IR (2003) Discovery of DNA polymerase. J Biol Chem 278:34733–34738. https://doi.org/10.1074/jbc.X300002200
21. Kornberg AR (2001) Remembering our teachers. J Biol Chem 276:3–11
22. The adhsgfhgd Figure 6. This work has been released into the public domain by its author, LadyofHats. This applies worldwide. LadyofHats grants anyone the right to use this work for any purpose, without any conditions, unless such conditions are required by law.
23. DNA Polymerase (2020). https://en.wikipedia.org/wiki/DNA_polymerase. Retrieved 21 July 2020
24. Chemistry LibreTexts (2017) DNA Replication, the Double Helix, and Protein Synthesis. https://chem.libretexts.org/Courses/Sacramento_City_College/SCC%3A_Chem_400_-General_Chemistry_I/Text/22%3A_Biochemistry/22.7%3A_DNA_Replication%2C_the_Double_

Helix%2C_and_Protein_Synthesis#: ~ :text=In%20DNA%20replication%2C%20each%20strand,synthesis%20of%20an%20RNA%20sequence. Retrieved 21 July 2020
25. Topoisomerase (2020). https://en.wikipedia.org/wiki/Topoisomerase. Retrieved 21 July 2020
26. Chien A, Edgar DB, Trela JM (1976) Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. J Bacteriol 127(3):550–1557. https://doi.org/10.1128/jb.127.3.1550-1557.1976
27. Saiki RK et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839):487–491. https://doi.org/10.1126/science.239.4839.487.PMID2448875
28. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230(4732):1350–1354. https://doi.org/10.1126/science.2999980
29. Theorell H (1959) The nobel prize in physiology or medicine 1959. https://www.nobelprize.org/prizes/medicine/1959/ceremony-speech/
30. HBB gene (2020) https://ghr.nlm.nih.gov/condition/sickle-cell-disease#inheritance. Retrieved 21 July 2020
31. Thalassemia. Down Syndrome. Cystic Fibrosis. https://online.regiscollege.edu/blog/information-5-common-genetic-disorders. Retrieved 21 July 2020
32. Cystic fibrosis. https://kidshealth.org/en/teens/cysticfibrosis.html#: ~ :text=Cystic%20fibrosis%20is%20an%20inherited,(cystic%20fibrosis%20transmembrane%20regulator). Retrieved 22 July 2020
33. Breast cancer. https://ghr.nlm.nih.gov/condition/breast-cancer#statistics. Retrieved 22 July 2020
34. Hamlin J, Sequin CH (2010) Computer generation of ribbed sculptures. J Math Arts 4:177–189. https://doi.org/10.1080/17513472.2010.487183
35. What is a gene? https://ghr.nlm.nih.gov/primer/basics/gene. Retrieved 20 July 2020
36. DNA/RNA chem.libretexts https://chem.libretexts.org/Courses/Sacramento_City_College/SCC%3A_Chem_400_-General_Chemistry_I/Text/22%3A_Biochemistry/22.7%3A_DNA_Replication%2C_the_Double_Helix%2C_and_Protein_Synthesis#: ~ :text=In%20DNA%20replication%2C%20each%20strand,synthesis%20of%20an%20RNA%20sequence. Retrieved 20 July 2020
37. mRNA in the nucleus. https://www.genome.gov/genetics-glossary/messenger-rna. Retrieved 23 July 2020
38. Transcription process. http://naturedocumentaries.org/10545/movie-rna-polymerase-ii-transcription-cheung-cramer-2012/. Retrieved 20 July 2020
39. Cramer P, Bushnell D, Kornberg R (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å ngstrom resolution. Science 292(5523):1863–1876. www.jstor.org/stable/3083926. Retrieved 25 July 2020. Image 1I50 creayed with ICM-Pro, MolsoftL.L.C.
40. Gnatt A, Cramer P, Fu J, Bushnell D, Kornberg R (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 angstroms resolution. Science 292 (5523):1876–1882. Image of 1I6H created with ICM-Pro, Molsoft LLC. www.jstor.org/stable/3083927. Retrieved 25 July 2020
41. Bernecky C, Grob P, Ebmeier CC, Nogales E, Taatjes D (2011) Molecular architecture of the human mediator-RNA polymerase II-TFIIF assembly. PLoS Biol 9(3):e10000603. https://doi.org/10.1371/journal.pbio.1000603
42. Lamy P, Fina F, Bascoul-Mollevi C, et al (2011) Quantification and clinical relevance of gene amplification at chromosome 17q12-q21 in human epidermal growth factor receptor 2-amplified breast cancers. Breast Cancer Res 13, R15:(12 pages). https://doi.org/10.1186/bcr2824
43. Genes, Medicine, Proteins. https://www.nigms.nih.gov/education/Booklets/the-new-genetics/Documents/Booklet-The-New-Genetics.pdf#page=66. Retrieved 16 July 2020

44. Palmer SN, Giesecke NM, Body, SC, Shernan SK, Fox AA, Collard CD (2005) Pharmacogenetics of anesthetic and analgesic agents. Anesthesiology. 102(3):663–671. https://doi.org/10.1097/00000542-200503000-00028
45. BCHE, butyrylcholinesterase. https://ghr.nlm.nih.gov/gene/BCHE#location. Retrieved 23 July 2020
46. Barbarino JM, Haidar CE, Klein TE, Altman RB (2014) PharmGKB summary: very important pharmacogene information for UGT1A1. Pharmacogenet Genom 24(3):177–183. https://doi.org/10.1097/FPC.0000000000000024
47. Dean L, Pratt VM, Scott SA, Pirmohamed M, et al (eds) (2012) Bethesda (MD): National Center for Biotechnology Information (US). Last Update: June 11, 2018. https://www.ncbi.nlm.nih.gov/books/NBK84174/. Retrieved 18 Feb 2021
48. Gorbalenya AE, Baker SC, Baric RS et al (2020) The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536–544. https://doi.org/10.1038/s41564-020-0695-z
49. Coronaviruses, Overview, Information for Researchers. https://www.niaid.nih.gov/diseases-conditions/coronaviruses. Retrieved 9 July 2020
50. Lou R, Zhao X, Li J et al (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 395(10224):565–574. https://doi.org/10.1016/S0140-6736(20)30251-8
51. Dai W, Zhang B, Jian, X-M, et al. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 Main Protease. Science. 368 (6497):1331–1335. Image of 6LZE created with ICM-Pro, Molsoft LLC. https://doi.org/10.1126/science.abb4489
52. St John, SE, Tomar S, Stauffer SR, Mesecar AD (2015) Targeting zoonotic viruses: structure-based inhibition of the 3C-like protease from bat coronavirus HKU4-The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS). Bioorg Med Chem 23: 6036–6048. Image of 4YOJ created with ICM-Pro, Molsoft LLC. https://doi.org/10.1016/j.bmc.2015.06.039
53. Diamond Coronavirus Science. https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.htm. Retrieved 9 July 2020
54. Hardy MA, Wright BA, Bachman JL, et al (2020) Treating a global health crisis with a dose of synthetic chemistry. ACS Cent Sci. https://doi.org/10.1021/acscentsci.0c00637
55. Corum J, Zimmer C (2020) Bad news wrapped in protein: inside the coronavirus genome. https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html. Retrieved 25 July 2020
56. Pollack A (2000) The next chapter in the book of life: structural genomics. https://archive.nytimes.com/www.nytimes.com/library/national/science/070400sci-genome-protein.html. Retrieved 25 July 2020
57. Zimmer C (2017) Scientists are designing artisanal proteins for your body. https://www.nytimes.com/2017/12/26/science/protein-design-david-baker.html. Retrieved 25 July 2020

Julie Ealy is an Associate Professor of chemistry at Penn State Lehigh Valley. She presently teaches general and organic chemistry. She has taught for 50 years, became interested in teaching at the university level, and graduated with a Ph.D. from Columbia University in 1998. Publishable scientific research has resulted with undergraduates, and she also pursues chemical education research. Outside of work, she spends her time in a log cabin home with her husband and their tan Airedale named Copper, whose sense of humor and naughtiness keep them busy. She has two beautiful granddaughters who live many hours away. One of the joys of teaching, I think/question/wonder about, is the need to repeat what I say to the 18-year old first-year student in college who only three short months ago was a senior in high school. As I have been heard to say after several repeats, “I love getting paid the big bucks to talk to myself.” Thank goodness for older age and a sense of personal humor.

The Art and Science of Biomimicry—Abstracting Design Principles from Nature

29

Laura Stevens, Deborah Bidwell, Michelle Fehler, and Asha Singhal

Come forth into the light of things, Let nature be your teacher.

~ William Wordsworth

Summary

Biomimicry is an emerging discipline that seeks nature's advice and brings diverse stakeholders together to create designs that emulate the way nature functions, not just the way it looks. The field itself is a multidisciplinary endeavor, yet biomimicry educators frequently work alone. Pedagogical methods based on trial and error may waste precious time. In this study, a set of four biomimicry experts from diverse disciplines and different areas around the globe collaborated to compare pedagogy and analyze student work to illuminate best principles for teaching students to translate biology into design solutions, a key step in the biomimicry design process. A total of 313 assignments created by 179 different students were evaluated. The results

L. Stevens (✉)
The Hague University of Applied Sciences, Hague, Netherlands
e-mail: l.l.stevens@hhs.nl

Delft University of Technology, Delft, Netherlands

D. Bidwell
College of Charleston, Charleston, USA
e-mail: bidwelld@cofc.edu

M. Fehler
Arizona State University, Tempe, USA
e-mail: mfehler@asu.edu

A. Singhal
Hybrid Futures, Berlin, Germany

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_29

showed that the inclusion of art in the learning of science, namely the hand drawing of the biological mechanism can lead to higher quality of abstracted design principles.

Art and Science of Biomimicry - artwork Jerlinga, Feltrinelli & Singhal

The code of this chapter is 01000010 01101001 01101111 01101101 01101001 01101101 01101001 01100011 01110010 01111001.

1 Introduction

Biomimicry is an emerging approach to innovative problem solving that looks to nature for resilient and sustainable solutions to human problems. The foundation of biomimicry lies within biology and is simultaneously ancient, but merges with engineering, design, and other disciplines such as chemistry, social innovation, and business. The adaptations and deep patterns present within the millions of species living on earth today represent 3.8 billion years' worth of time-tested, sustainable solutions to the same challenges that humans now face. Humans are (re)learning to apply these functional biology design lessons through the process of biomimicry design thinking. Practitioners of biomimicry do not just learn about nature, they learn from nature [1]. The Biomimicry Institute states that biomimicry aims to realign human actions with nature's principles, to promote a viable, equitable, and livable world.

The demand for nature-oriented design education and improvement of twenty-first-century teaching has grown exponentially in recent years. The scope of biomimicry has been explored in depth by pioneers such as Janine Benyus in her book *Biomimicry—Innovation Inspired by Nature*, and by Dayna Baumeister in her book *Biomimicry Resource Handbook*. Yet, engaging in the process of biomimicry thinking takes practice and an initial investment of time. Over the past 10 years, biomimicry has found its way into kindergarten workshops, K-12 programs around the globe to an entire Master of Science in Biomimicry program, initiated by Baumeister in 2015 [2]. On the other hand, nature-driven designs didactics have only just begun to take root.

Biomimicry can be distinguished from bio-inspired design in that biomimetic designs are held to a rigorous scientific standard. Living systems are made up of complex, ever-changing networks of interdependent organisms surviving within this complexity. When biomimics look for biological models, we look for adaptations that have been honed to solve for specific problems. Biomimicry asks nature's advice by focusing on what we need our design to do in a specific context and matching that function and context to natural mentors that have solved the same functional challenge in beautiful, elegant, efficient, regenerative, and resilient ways. What may be most striking about biomimetic design is the aesthetic beauty with which design challenges are solved, and just as in nature, aesthetics can be highly adaptive in design.

What is the art behind the science of biomimicry? How are the biological strategies and mechanisms from adaptations in nature translated into abstracted design principles for biomimicry practitioners, innovators, architects, and designers? This chapter reviews the steps necessary to teach that translation, highlights the importance of drawing to learn, and showcases artful illustrations made by student biomimicry practitioners during this process. Preliminary research showed that when teaching biomimicry students how to bridge the gap between biology and design through abstracting design principles, the lack of scientific research on pedagogical methodologies leaves teachers and students exploring the design process together through trial and error. Observations and student survey responses in the study 'Analogies in Biomimicry' [3] showed that students continually struggle with understanding how to transfer biology to design. However, in this same study, Stevens found that first drawing the biology and subsequently drawing the abstraction of how that biology is functioning, aided in her students understanding the design principles and implementing these into design solutions. Is this true elsewhere? Biomimicry is a team effort, and successful biomimics rarely work alone. Yet biomimicry educators frequently do. This chapter brings four biomimicry educators (three professors and a biomimicry practitioner) together to examine and share best teaching principles in the field for the translation phase between biology and design, evaluating this phase through the work of 205 students. The authors share a background of learning biomimicry from the same Master of Science program at Arizona State University (ASU).

2 Framework

2.1 Biomimicry Thinking

Biomimicry design thinking is the methodology which merges biomimicry thinking and design thinking, following the design phases of scoping the challenge, discovering existing solutions, creating ideas, and evaluating them to create innovative design solutions [3]. Within biomimicry, each phase is focused on what we, as designers, can learn from nature. In the scoping phase, biomimics ask what we want our design to *do,* identifying its function and biologizing that function so that we can ask nature's advice about how to solve it. Just as a well-adapted organism thrives in natural systems by optimizing its fitness and adapting to an ever-changing context within its unique niche, a well-adapted human design must meet the functional needs of the design challenge within the context it operates in order to optimize its success [4]. When biologizing the design question, biomimics ask, 'How might nature do ____?' [5]. For example, how might nature regulate temperature? The discovering phase of biomimicry then looks to natural organisms, processes, or ecosystems that have evolved adaptations for solving the same functional design problem and context we face, as opposed to looking for inspiration from previous designs made by other human designers, which is often the norm in ideating phases of the design thinking process [6]. The goal of biomimetic design is to learn from the research and development throughout which life has survived, applying scientific principles to create resilient human innovations that are vetted by science and sustainable by design.

ASU teaches that researching the *function, strategy, and mechanism* of adaptations is key for translating those from nature into human design. For example, a cactus excels at 'regulating temperature.' Biomimics explain the natural strategy, or 'how the cactus regulates temperature.' The strategy of the cactus might be described as 'by creating shadow and air flows.' The next step is to consult scientific literature to determine how the natural mechanism of this strategy works. For example, research articles describe detailed mechanisms regarding how much shade is provided by ribs and spines and more specifically how the shadows and air flows are created (Fig. 1).

During desktop research to find organisms fitting their design need, biomimicry design students iterate to become adept with the function–strategy–mechanism paradigm. In the discovery design step, student teams explore dozens of possible natural mentors to describe their function and strategies before honing in on those which seem to carry out the function best. A deep dive into scientific articles locates

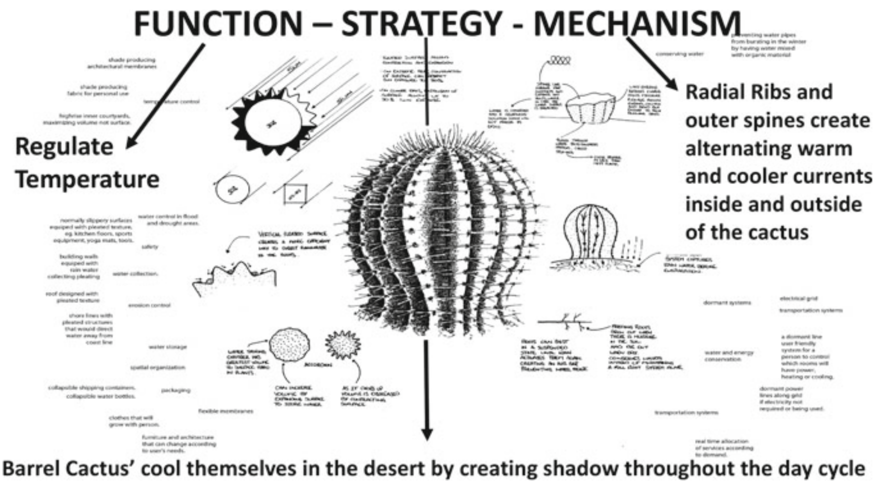


Fig. 1 Example of function–strategy–mechanism presentation published with permission by Rui Felix (adjusted with biomimicry terms)

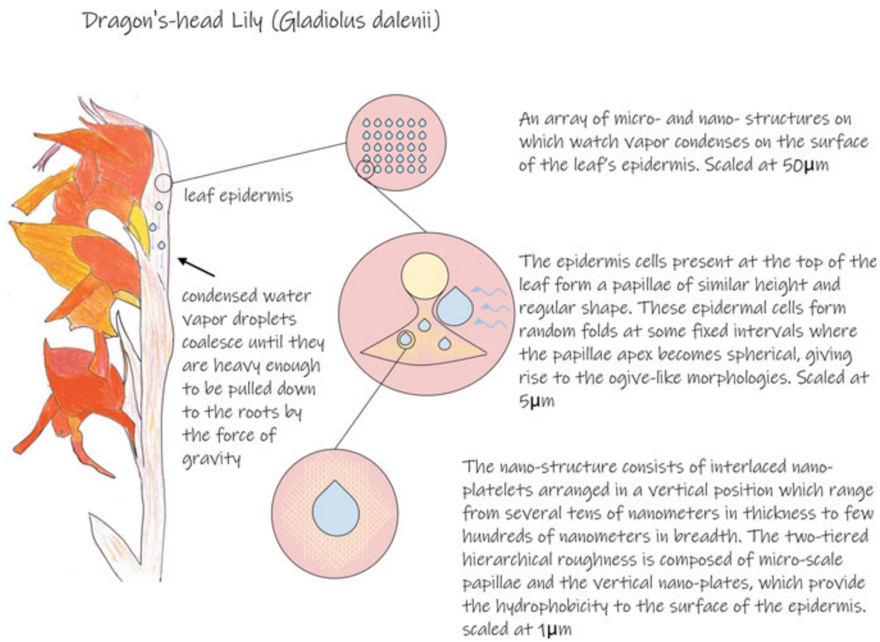


Fig. 2 Biological design principle and diagram of *Gladiolus dalenii,* C.M. Langford

primary research to reveal detailed mechanisms behind nature's elegant solutions. Each organism is recorded in a 'bio-brainstorm' collection spreadsheet. This includes the function, organism's common and scientific name, its strategy and mechanism fulfilling the function, quoted literature citations, and cited references.

Subsequently, biomimicry students hone in on and summarize promising mechanisms as biological design principles (Fig. 2) describing both in text and visually how the biological function is carried out. Finally, they reach the crucial and most difficult step, writing an abstracted design principle in which biological terms are eliminated and exchanged for non-biological terms to make this natural design lesson accessible to designers and engineers (Fig. 3).

Abstracted design principles (ADPs) are illustrated, demonstrating how the mechanism works including the scale, texture, subcomponents, behavior, or interconnecting functions. The goal is to use the ADP diagrams as a starting point for design ideation [4]. A Nature's Technology Summary (NTS) is the template we assign for capturing and summarizing the research and ADPs for each of the biological mentors that make the final cut.

Analogical thinking or reasoning is the theory of using what is learned in one context (such as biology) and applying it to a second context, such as design [7]. Within the field of biomimicry, three distinct levels of hierarchy in this analogical thinking are described from the most simple to the most intricate: (1) form analogies, or those emulating the functional shape or structure of an organism;

ABSTRACTED DESIGN PRINCIPLE

An erected unit coated in micro-protrusions that fold at fixed intervals form spherical apexes with concave bases at heights of 10.2 ± 1.5 μm, diameters of 5.5 ± 0.4 μm and cavity depths of 6.2 ± 0.5 μm. The apexes are hierarchically arranged to form rows and columns with longitudinal spacing of 18.3 ± 2.5 μm and transversal spacing of 36.6 ± 2.9 μm. The fleshy protrusions have vertically positioned, interlaced platelet nano-structures on their surfaces. The micro and nano-structures have surface contact angles of 135° ± 0.4°. the coating condenses water vapor at the protrusions base, with secondary condensation sites on the surface nano-structures. Droplets grow in size and eventually gravity pulls the droplets downward.

ABSTRACTED DESIGN PRINCIPLE ILLUSTRATION

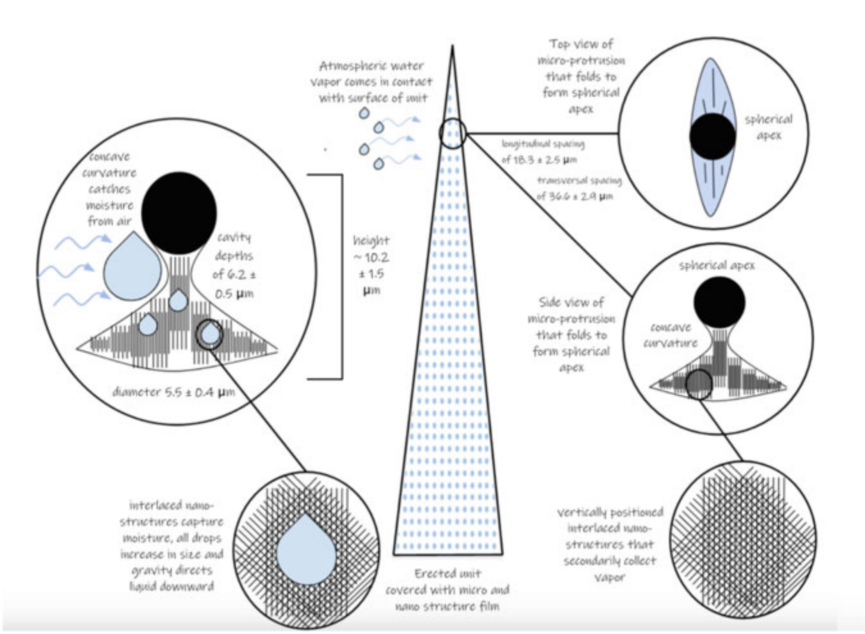


Fig. 3 Abstracted design principle and diagram of *Gladiolus dalenii*, C.M. Langford

(2) process analogies, or those emulating the functional behavior or biochemistry of an organism; and (3) system analogies, when an entire ecosystem's function is emulated [4]. Mimicking systems is considered the highest level of analogical thinking within biomimicry design.

In this study, we illustrate, describe, and analyze biomimicry students' attempts to write and illustrate ADPs within their NTSs to evaluate effective and ineffective methodology to illuminate current best pedagogical principles.

2.2 Drawing in Science

One method for improving understanding of complex ideas is through drawing scientific visualizations which are important learning aids in science education [8].

Developmentally, drawing is motivational, and engaging for children and drawing to show understanding in the higher education classroom also captivates student interest and may amplify creative thinking by stimulating interactions between brain hemispheres [9, 10]. Cognitive functions such as observation, problem solving, explanation, and communication, which form the foundations of scientific thinking, interrelate with the basic mechanisms of illustrating [11]. Drawings are fundamentally critical portrayals that incorporate a few subtleties and preclude others; subsequently, every impression a student makes in a drawing might be reflective of what students are and are not learning [12]. Drawing enables students to synthesize ideas, training them to identify any gaps in their understanding [13–15]. Biomimics must accurately understand biological adaptations to create designs that function the same way that nature does. Although it is important for biomimicry students to be explicit in the visual representation of the natural adaptive mechanism, only basic visual communication skills are required to promote understanding, the process is more important to the learning than the outcome [16, 17]. Drawing also facilitates communication of the mechanism to other audiences, assisting with the translation of the functional mechanism found in nature into engineering lexicon. Drawing with accuracy enhances observation skills and learning. There is a feedback mechanism involved, the act of drawing both requires and enhances understanding [18].

Drawing in science serves various purposes in student learning. When looking through the lens of abstracting design principles in biomimicry, drawings serve multiple pedagogical goals: drawing as a mode of internalizing science, drawing to enhance observation, drawing to enhance model-based reasoning, drawing for problem solving, drawing to connect concepts/ideas, drawing to learn, drawing to reveal student's mental models, and drawing to communicate [13]. Drawing can help make the 'unseen seen'. To understand forms, processes, and systems in nature, drawing requires a deep understanding of the details. It allows an external process to internalize information [13] and can visually pinpoint where a gap is in understanding. Biomimicry can only emulate nature successfully if it follows what nature does in an authentic way. This forces the biomimic to look closely, investigate the natural model from different perspectives, and explore specifically how the adaptation functions [19]. Even biologists are asked to draw abstract visual models to support reasoning while solving problems based on complex concepts [13].

Student drawings in biology can range across broad spectra. They can vary across scales from atomic to ecological; they can differ in their incorporation of text, e.g., flowcharts are created predominantly with words, indicating relationships. On the other end of the spectrum, drawings depicting species morphology might contain few to no labels and can range from detailed to abstract depending upon the context. While a thoroughly detailed drawing might be useful for understanding bird behavior, only a simple shape with the word bird inside it will be sufficient for an ecosystemic concept map outlining interspecific relationships [13].

3 Research Question

Although biomimicry education is expanding exponentially, teachers and students still struggle with getting the science accurately and visually communicated into design principles that can be used for innovative ideas. The authors have the same foundational biomimicry education, but teaching at different schools to different student audiences results in variations on our desired outcome. What are these differences and how can we rigorously funnel what we have learned through iterative curriculum development into recommended pedagogical principles? Our research question is: Which characteristics, methods, factors, descriptors, learning outcomes, or techniques are most often present in biomimicry student work that correlates to the highest-quality abstracted design principles?

3.1 Pinpointed Sub-Questions

- What subcomponents of the biomimicry thinking methodology are most vital for students to achieve high-quality abstracted design principles?
- What curricular or pedagogical factors influence whether students achieve systems-level abstracted design principles?
- What is the art behind the science of biomimicry? Does drawing to learn improve the learning outcomes of multidisciplinary novice biomimicry practitioners? If so, how is it correlated with achieving high-quality abstracted design principles?

Our research focused on the quality of the Nature's Technology Summary (NTS) to normalize our diverse set of data. We want to find out how each of the methods from our classrooms improves the communicative starting point for biomimicry design for interdisciplinary design teams, namely the abstracted (biological) design principle. Does the primary research offer explanatory diagrams with scaled measurements? Are there visuals of how a mechanism moves? Have students observed details while drawing what they see in nature? These are a few predictions of what may arise from this research. By exploring these factors, we aim to identify which methodological steps are essential to reach a comprehensive and useful visual translation and, when these factors are defined, how to introduce more complex translations to apply to design.

4 Method

4.1 Context and Participants

In this study, we analyze a single biomimicry assignment, the NTS, given across four separate university student cohorts over a two-year period using both quantitative and qualitative approaches to improve our result validity [20]. We identify recommended pedagogical principles for teaching and learning the crucial and most challenging step of the biomimicry design process: abstracting the design principle. Student populations varied between undergraduate and graduate levels and ranged across a variety of disciplines, allowing for the comparison between design students and interdisciplinary students lacking design backgrounds and between novice first-year undergraduates (a control group) and experienced upper-level undergraduate and graduate student populations. Universities included in the study are The Hague University of Applied Sciences (THUAS), Arizona State University (ASU), and the College of Charleston (CofC). A total of 313 NTS assignments created by 179 different students were evaluated. Our preliminary recommended pedagogical principles findings were applied to the most recent test group (ASU, summer 2020). See Table 1.

5 Description of Common Assignment

The common assignment for all cohort participants, the Nature's Technology Summary (NTS), is described in detail below. Students at THUAS, ASU, and CofC all completed the NTS assignment as a major element in the semester. A NTS template was provided to all student populations. Students followed the same six-step process. To prepare for the NTS assignment and engage in observations of biology throughout the courses, a variety of preliminary observational, research, and drawing tasks were assigned (Fig. 4a–c).

5.1 Assignment Steps

Step 1 *Bio-Brainstorm*: This pre-NTS step includes seeking organisms that are successfully solving the same challenge in the same context as the human design challenge. Students conduct preliminary research including a broad spectrum of potential organisms from wide ranging taxa and scales, representing diverse strategies. Honing on in their top choices, students go deeper to thoroughly investigate the primary literature. The research on these top choice organisms is summarized using the NTS template (Fig. 5a–c).

Step 2 *Function–Strategy–Mechanism*: Students write a short natural history, define one biological function, write out the strategy that allows the organism to meet this function, and then write a longer piece about the specific biological mechanism that achieves this strategy. Direct primary literature citations should be included, complete with in-text and final references (also shown in Fig. 5a–c).

Table 1 Research context, participants, and other details

Institution	The Hague University of Applied Sciences (THUAS)	Arizona State University (ASU)	College of Charleston (CofC)
Location	The Hague, Netherlands	Tempe, Arizona, USA	Charleston, South Carolina, USA
Audience	Design, engineering, other miscellaneous technical fields	Architecture, industrial design, interior arch., landscape arch., visual communication design, sustainability	Biology, entrepreneurship, urban studies, environmental and sustainability science, first-year experience
Level	Undergraduate	Undergraduate and graduate	Undergraduate
Cohort dates	Spring 2019 Spring 2020	Fall 2019 *Summer 2020	Spring 2019 Fall 2019 Spring 2020
Number of participants and number of NTS' made	2019 n = 22 (16 NTS) 2020 n = 23 (35 NTS)	2019 n = 15 (30 NTS) *2020 n = 12 (12 NTS)	Sp2019 n = 40 (112 NTS) F2019 n = 20 (19 NTS) Sp2020 n = 47 (89 NTS)
Student background	Minor for exchange students (motivation letter) or fourth semester for IDE students	Graduate students from various design disciplines (ARCH, IND, INT, ALA, VCD) as well as students from sustainability	Variable. No prior design experience. Upper level tends to have more biology, sustainability, and environmental science background
Course name(s)	Design with nature, industrial design engineering semester	Biomimicry in design	Special topics: biomimicry thinking and biomimicry, nature as mentor

*Summer cohort ASU = special test group with preliminary 'recommended pedagogical principles' findings integrated in instruction

Step 3 *Drawing to Learn*: Students (hand) draw the biological function, strategy, and mechanism as is described in the scientific research and describe the principle of what is happening in a succinct manner. This 'drawing to learn' process creates a biological mechanism drawing that visually communicates how the organism achieves its function (Fig. 6a–c).

Step 4 *Abstracting the Design Principle (ADP)*: Students synthesize and describe the causal mechanism replacing the biological terms with non-biological (i.e., engineering) terms to identify the lesson that natural organisms have for human designers, the abstracted design principle (ADP) (Fig. 7).

Step 5 Step 5 *Drawing to Communicate*: Students visually communicate the adapted design principle to designers and engineers note, CofC students

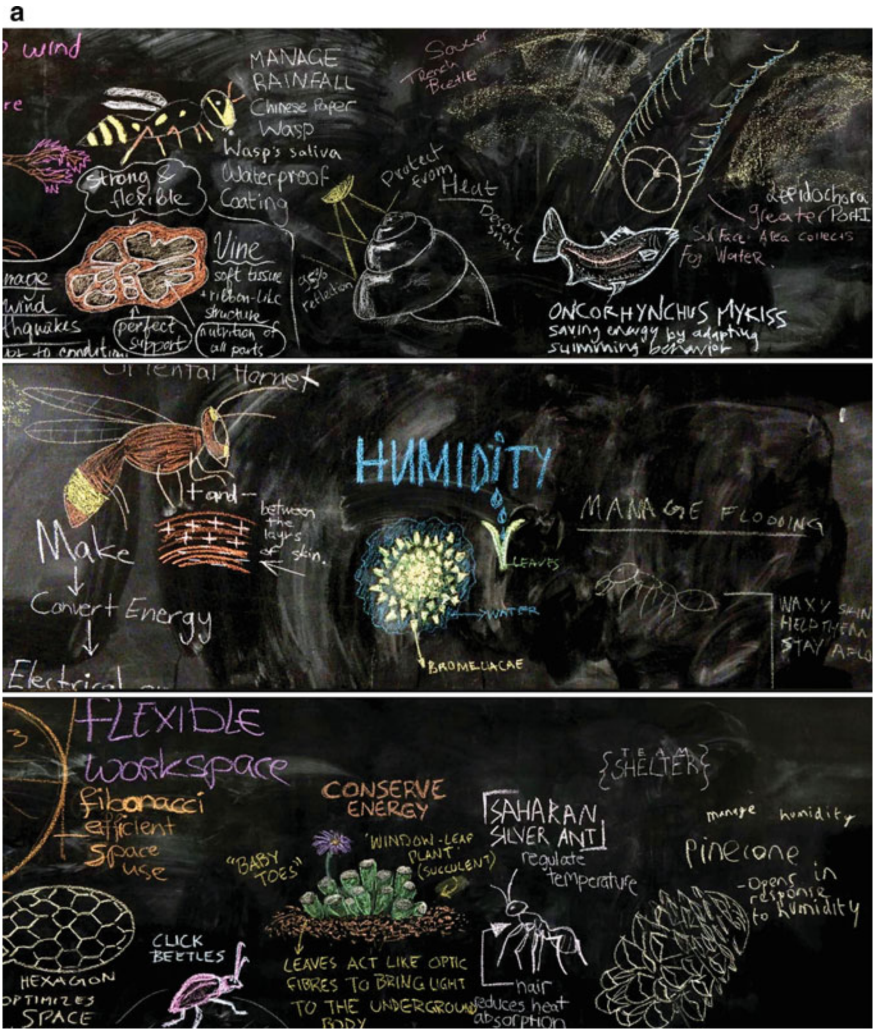


Fig. 4 **a** Pre-NTS bio-brainstorming and drawing to learn, THUAS students spring 2020. **b** iSites drawing to learn how to observe and see, H. Carter, ASU student summer 2020. **c** iSites drawing to (re)connect with local flora, describe biological adaptations, and learn how to observe and see, E. Peters, CofC student spring 2019

lack design experience, and although hand drawing was strongly emphasized as a teaching and learning tool, they were given the option to include either hand drawings or well-chosen images from the primary literature in their final NTS submissions (Fig. 8a–c).

Step 6 *Brainstorming and Emulation*: The design phase where students ideate and iterate biomimetic design ideas is not part of this study, but is shown here (Fig. 9a–c) to aid the reader's understanding of how ADPs are used.

b

Pricklypear (Opuntia basilaris)

5.23.20
3:47 pm
Millett Ranch
Its 90 feels 85 degrees with little to no clouds with little wind that is blowing northwest.
The sun is bright but not too hot for being outside. The birds are singing loud today, and I wish I could capture the purple in the prickly pear cactus.

Notes: This cactus may consist of hundreds of pads that are green, and blue-gray color, and can grow five inches long. During summer, the pads shrivel up because they pump out the water they stored during winter.

https://calscape.org/loc-california/Opuntia%20basilaris()

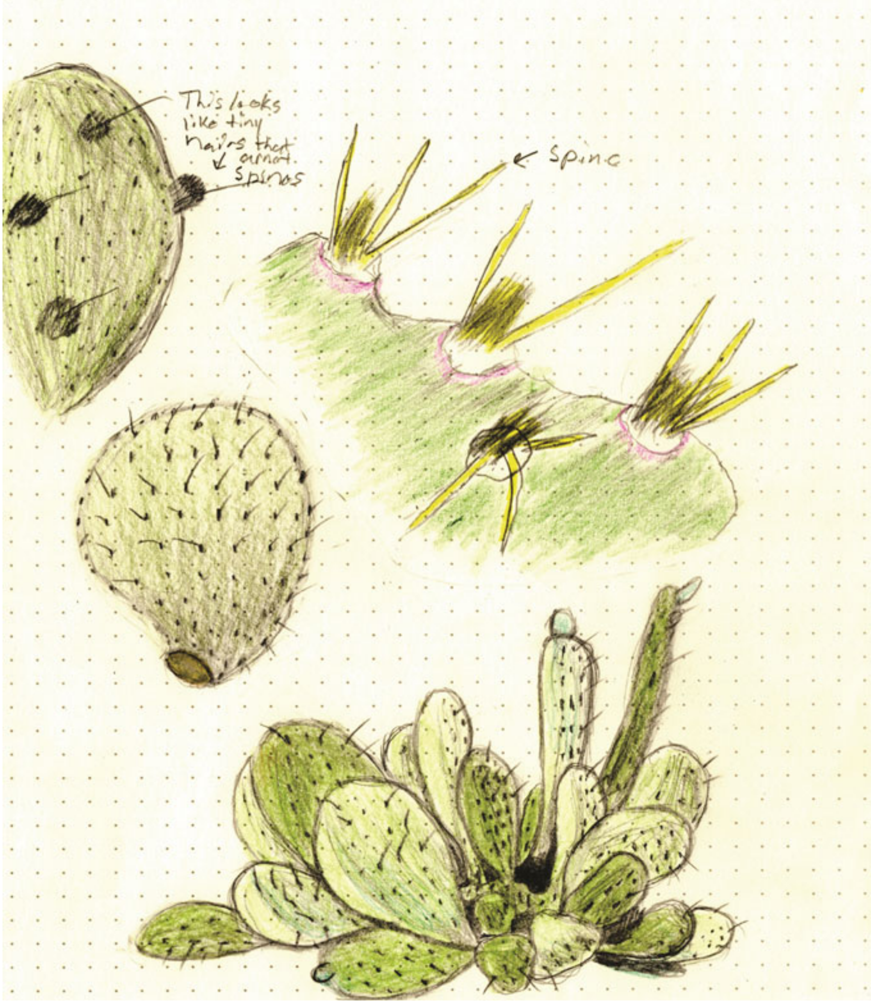


Fig. 4 (continued)

c

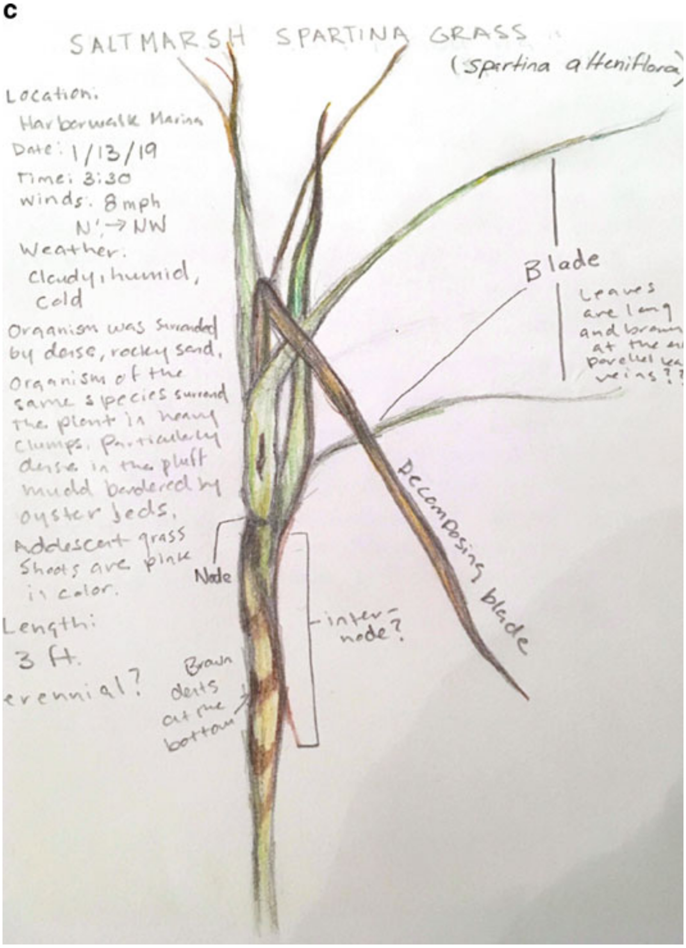


Fig. 4 (continued)

5.2 Distinction Between Student Cohorts

5.2.1 THUAS

Students from The Hague University of Applied Science conducted a Challenge to Biology (C2b) design solution project using the biomimicry design methods [4] combined with design thinking, from scoping through evaluation to incorporate methodologies into a design-based learning (DBL) project. The non-obligatory NTS assignment at THUAS culminated the discovery phase and prepared students for the creative brainstorming process. Sharing iterative drawing to learn exercises by hand was a main factor as well as technical drawings of how the mechanisms worked within a final prototype.

a

NATURE'S TECHNOLOGY SUMMARY | (DESIGN BRIEF APPENDIX)

Contact: Agostina Feltrinelli

Natural History:

The western honey bee is native to Europe, the Middle East, and Africa, and, though none of its at least 20 recognized subspecies naturally occur in the Americas, these have been extended beyond their natural range due to the economic benefits brought about by they pollination and honey production. Because of this, honey bees are now naturalized on all continents apart from Antarctica.

The non-reproductive females, the worker honey bees, are the smallest in size, with their bodies specialised for pollen and nectar collection, as their hind legs have a corbicula (pollen basket) which is specially designed to carry large quantities of pollen back to the colony. They also produce wax scaled on the underside of their abdomen, which are used to construct the wax comb within the colony, and have a barbed stinger with a poison sac which is torn from the end of their abdomen when they sting a tough-skinned victim such as a human. The act of stinging, a defensive behaviour worker bees use to protect the colony, results in the bee's death.

Honey bees have a highly social life history and their colonies could be considered to be superorganisms, with the entire colony being viewed as a biological unit. Their reproductive process, or swarming, is based on the premise of producing more colonies rather than individual bees.

Honey bees typically swarm in spring and early summer, when pollen and nectar are plentiful (Mortensen, Schmehl, and Ellis, 2013).

Function:	Modify chemical/electrical state, electric charge; Maintain community, provide ecosystem services through pollination; Get, store or distribute resources, capture solids; Get, store or distribute resources, distribute solids.
Strategy:	Honey bees collect and release pollen through the hairs on their bodies, whose static charge varies from that of the pollen.
Champion:	Honey Bee *Apis Mellifera*

Description Text:

Under clear, fair-day conditions, plants generally have small negative surface charges and are therefore surrounded by low intensity electric fields. The magnitude of their electric fields depends on the chemical composition of the plants, its height, and the environment, as, under unstable weather conditions, these fields can change polarity, with their surface charges becoming positive. The distribution of its electric field, on the other hand, varies with its shape, with sharp points such as flowers exhibiting greater electrical fields.

In contrast, foraging bees, such as honey bees, usually have electrically positive surface charges which occur when they fly through the air. In this manner, when a bee flies through the air, it is confronted with electrical currents and its body is electrostatically charged with frictional electricity. Research has suggested that pollen seeking insects' ability to accumulate pollen on their surface and later distribute it is enhanced by the forces of attraction between the insect's

Fig. 5 **a** Example steps 1 and 2 NTS (organism, function, and context), A. Feltrinelli, THUAS, spring 2019. **b** Example steps 1 and 2 NTS (organism, function, and context), K. Boakye, ASU, summer 2020. **c** Example steps 1 and 2 NTS (organism, function, and context), M. Gonzales, CofC, spring 2020

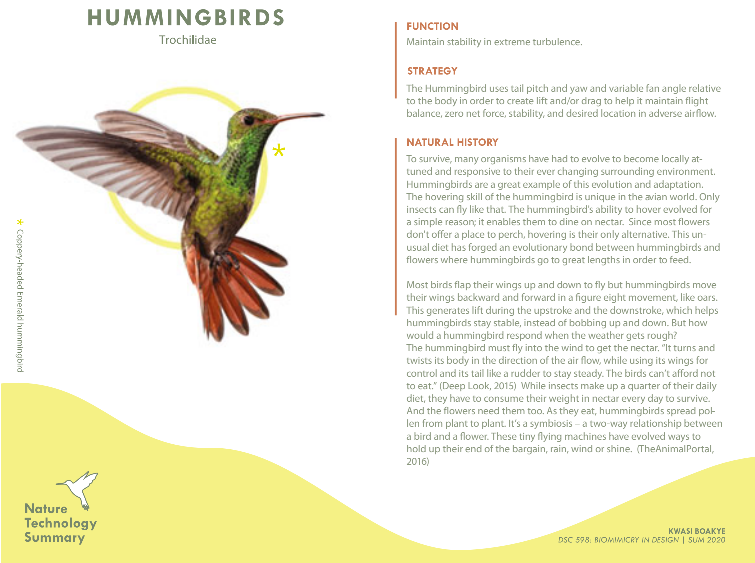


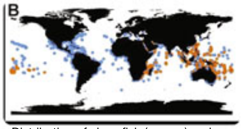
Fig. 5 (continued)

5.2.2 ASU

Students from the ASU summer 2020 course built their NTS based on a biology to design (B2d) process rather than the more common challenge to biology (C2B) process because of the shorter 5-week period. Starting with the organism, process or system of choice, they dove deep to discover just one strategy to translate into a design principle, focusing on one function. The ASU NTS template also asks students to specify at least four life's principles that the organism is exhibiting to recognize the deep patterns and connections in nature. *This test group integrated our preliminary 'recommended pedagogical principles' findings into the NTS instructions.

5.2.3 CofC

Students from the College of Charleston (CofC) First-Year Seminar (FYSE) and upper-level courses conducted a month-long challenge to biology (C2B) design project using the biomimicry design method [4] from scoping through evaluation. The NTS assignment at CofC concluded the discovery phase and prepared students for the creative brainstorming process. While it was emphasized and strongly encouraged to hand draw both the biological and functional technology mechanisms employing 'drawing to learn' methodology to formulate abstracted design principles, hand drawn visuals were not mandatory for the final NTS submission. Students in the CofC cohorts came from a diverse spectrum of disciplines including biology, entrepreneurship, urban studies, and environmental and sustainability



Distribution of clownfish (orange) and sea anemone (blue) across the globe (Litsios et al., 2012)

NATURES TECHNOLOGY SUMMARY (NTS)

Your Name: Mylene Gonzales

SUMMARY

Champion Mentor (common & Latin name)	Clown anemonefish (*Amphiprion ocellaris*) and Sea Anemone (*Actiniara*)
Organisms natural history context (habitat & conditions driving the evolution of its strategy)	Clown anemonefish are a species of damselfish that live in the warm waters of the Indian and Pacific oceans (users.on.net). Similarly, sea anemone are found in the tropical seas of the Pacific, but can also be found in coastal and temperate waters. (Tolweb.org). Mutualism between clownfish and sea anemone is well-known and is one of many mutualistic relationships that have evolved in coral reef ecosystems. Because clownfish are poor swimmers, they must hide in the tentacles of sea anemones (which have stinging nematocysts) in order to escape from their predators like larger fish, eel, and/or sharks (Mebs, 2009). Sea anemone's also benefit from this interaction because the clownfish scare away predators that usually eat anemone (Godwin and Fautin, 1992) and the excretory waste of clownfish also provide the anemone's zooxanthellae with nutrients.
Function (What does the technology do? Pasted from the FBS template)	The mutualistic relationship is an example of maintain community, cooperate between different species from the biomimicry taxonomy. Mutualism also increases the fitness of both individuals (Porat and Chadwick-Forman, 2004)
Mechanism (how the strategy works pasted from FBS worksheet)	The clownfish coats itself in the mucus of the anemone in order to protect itself from the stinging nematocysts. Clownfish is able to inhabit the anemone and is protected from other fish that may try to eat it through the stinging tentacles (Mebs, 2009). The sea anemone gets protection from its own predators by the territorial aggression of the clownfish that fight off other fish that may swim close to the anemone (Godwin and Fautin, 1992). It also gets nutrients from the clownfish when the clownfish excretes undigested waste and it falls onto the symbiotic dinoflagellates of the anemones that live in its tentacles.
Design Principle (idea that can be emulated in design, pasted from the FBS worksheet)	Daily close-knit interactions between two dissimilar individuals can foster meaningful relationships. Overtime, both individuals gain each other's trust, support, and other benefits that allow each individual to achieve positive personal growth.
Lifes Principles (use both headings and subheadings from the LP sheet and pick 2-4 that best fit)	Evolve to survive- replicate strategies that work Be locally attuned and responsive- cultivate cooperative relationships Be resource efficient- recycle all materials

Fig. 5 (continued)

studies, and had little to no prior design experience. Learning outcomes for the first-year experience students emphasized research methods, creativity, growth mindset, and communication skills. Upper-level students (spring) were almost exclusively junior and seniors (third and fourth year), as such, learning outcomes and expectations were elevated.

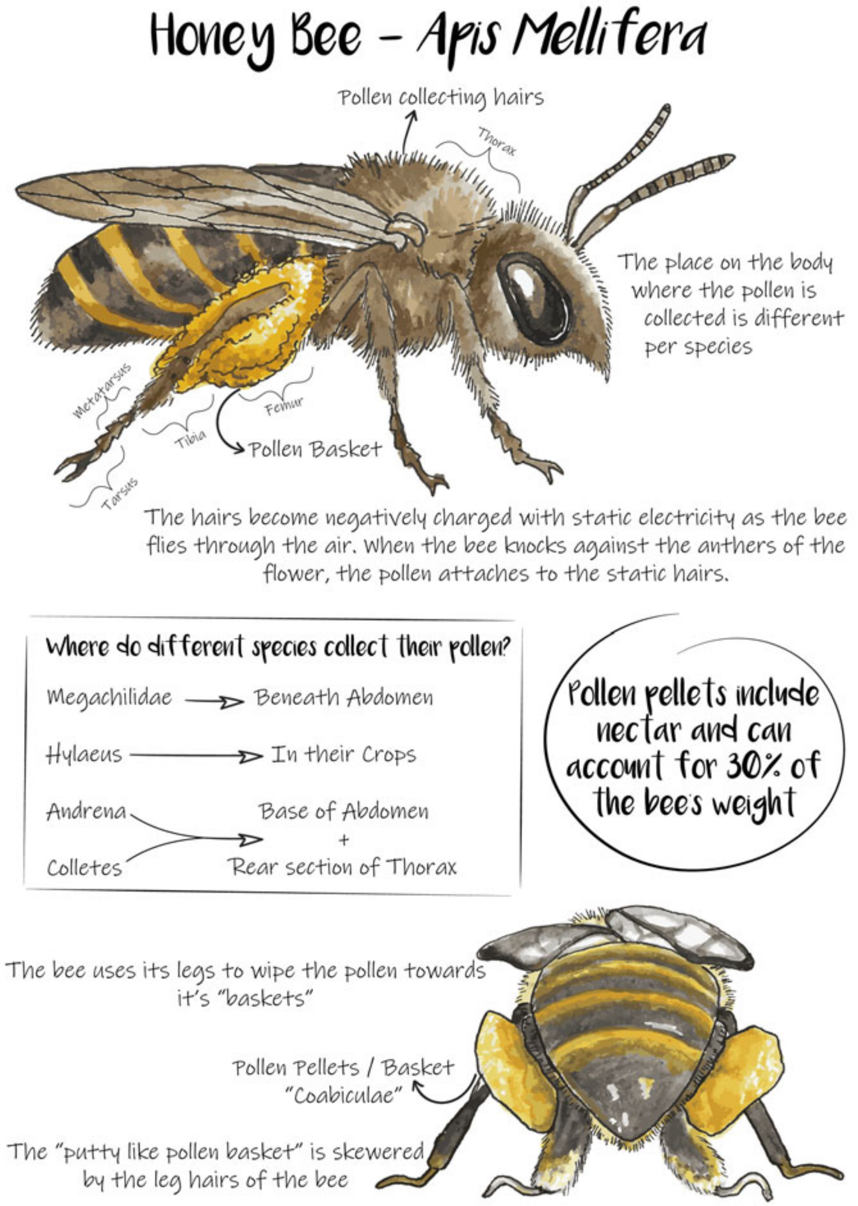


Fig. 6 **a** Step 3 NTS (BDP), E. De Bloois; A. Feltrinelli, THUAS. **b** Step 3 NTS (BDP), K. Boakye, ASU. **c** Step 3 NTS (BDP), M. Gonzales, CofC

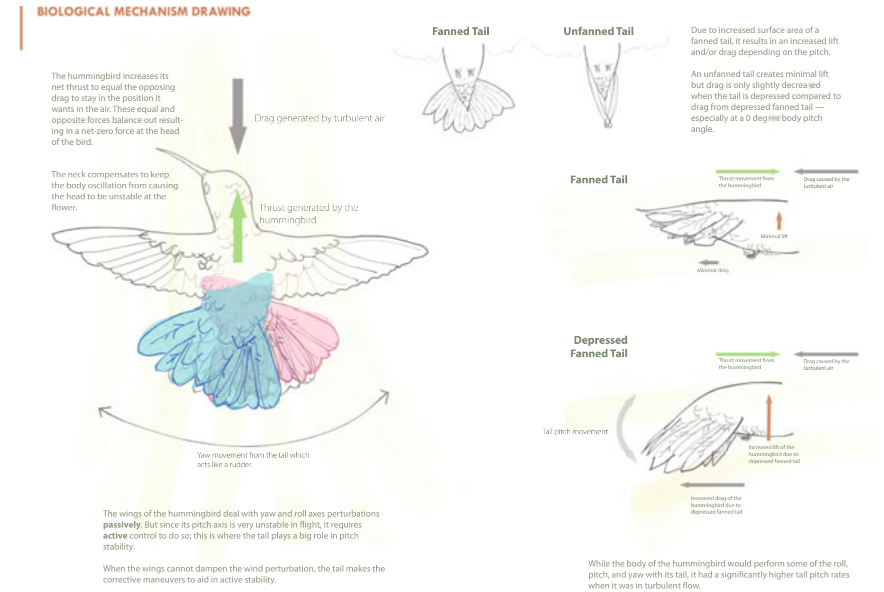


Fig. 6 (continued)

6 Data Collection and Analysis

6.1 Phase 1: Assignment Cues Comparison

Syllabi, assignments, and lesson plans were shared between author faculty at three participating universities (THUAS, ASU, and CofC). We compared our teaching methodology. For this study, student Nature Technology Summaries (NTS's) deliverables were collected and shared between the authors after Institutional Review Board (IRB) approval and/or student consent for publication. The same external assessor scored each of the student NTS assignments with an identical rubric recording categorical Y/N data about the clarity and depth of student work.

6.2 Phase 2: Rubric Analysis of NTS's

A scoring of Yes = green and No = red made visible whether an NTS element was included or lacking (Figs. 10 and 11). These empirical data made statistical comparisons possible. We investigated which elements aid in reaching accurate abstracted design principles (ADPs) that are both true to nature and beneficial to designers as they start the ideation phase. We also recorded whether ADPs were form, process, or system based. Systems-level analogies were emphasized within

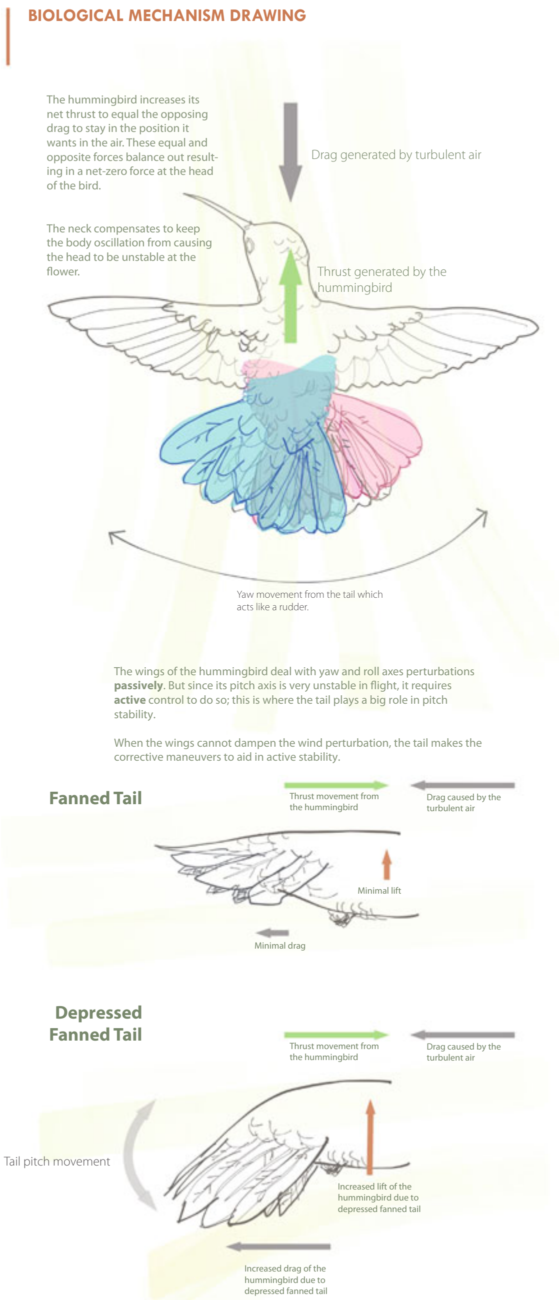


Fig. 6 (continued)

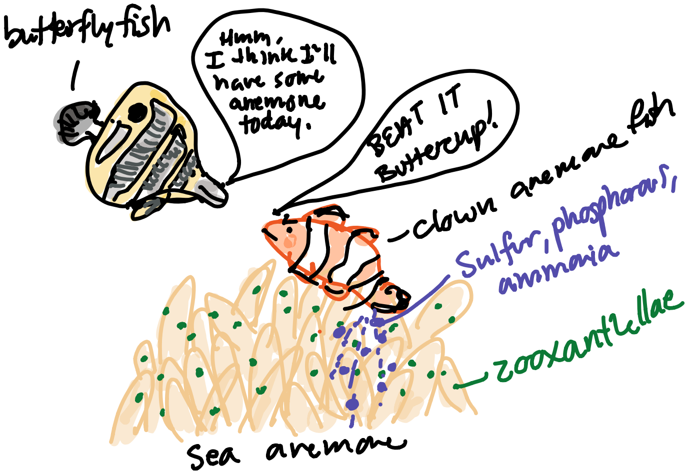


Fig. 2. Sea anemone gets benefits of this mutualism by 1) clownfish defending the anemone from its own predators like butterflyfish 2) clownfish secretes undigested waste (sulfur, phosphorous, ammonia) that become food for symbiotic zooxanthellae living inside the cells of sea anemone (Mebs, 2009)

Fig. 6 (continued)

"Fibres develop a positive static charge due to external factors such as motion, whereas separate modules have a negative charge, making them ideal to stick to these fibres. Due to their opposite charges, when the modules come into contact with the fibres, the modules stick to the fibres.". a) ADP Honey bee (*Apis Mellifera*), A. Feltrinelli, THUAS spring 2019
"The design uses an expandable flap that can pitch, roll, and yaw opposite the direction of adverse airflow to generate an increase/decrease in lift and/or drag in order to help reduce energy use, increase stability, and/ or support continuous maneuverability in turbulent airflow to stay in desired location". b) ADP Hummingbird *(Trochilidae), K. Boakye, ASU summer 2020*
"Daily close-knit interactions between two dissimilar individuals can foster meaningful relationships. Overtime, both individuals gain each other's trust, support, and other benefits that allow each individual to achieve positive personal growth" c) ADP Clown anemonefish *(Amphiprion ocellaris)* & Anemone *(Actiniara), M. Gonzales CofC spring 2020*

Fig. 7 Step 4 of NTS (ADP text): THUAS (**a**), ASU (**b**) & CofC (**c**)

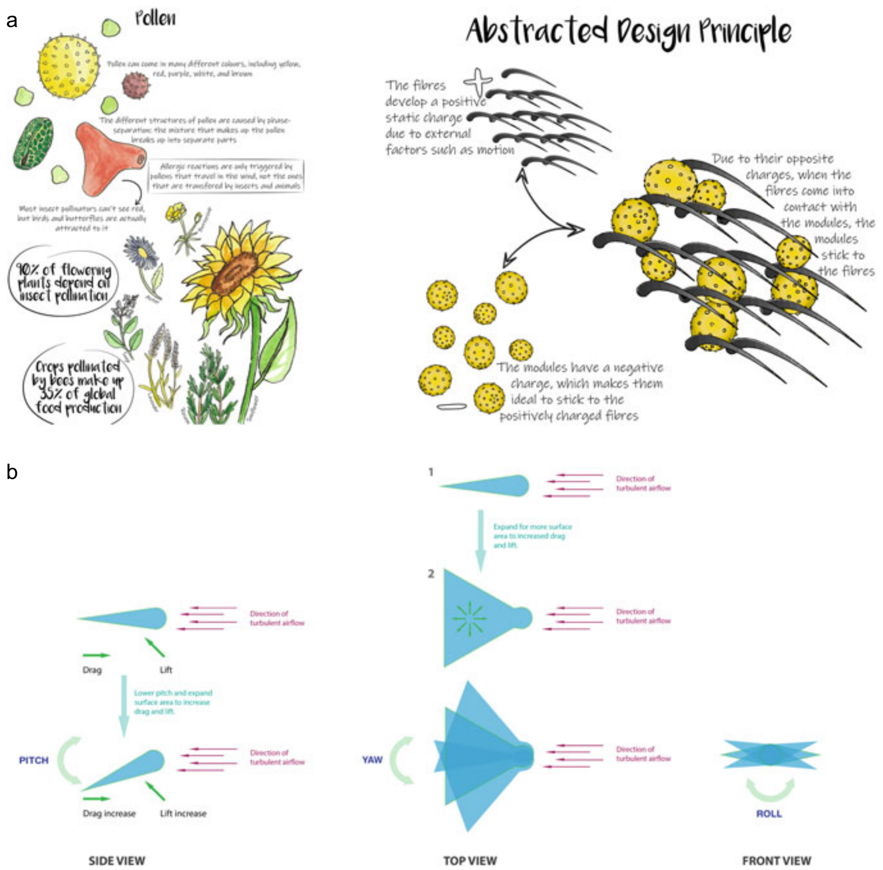


Fig. 8 **a** NTS Step 5 ADPs from honey bee (Apis Mellifera), E. De Bloois; A. Feltrinelli, THUAS. **b** NTS Step 5 ADPs from hummingbird (*Trochilidae*), K. Boakye, ASU. **c** NTS Step 5 ADPs from clown anemonefish (*Amphiprion ocellaris*) and anemone (*Actiniara*), M. Gonzales CofC

THUAS and CofC courses in 2020. To rank as a high-quality mechanism (*Y*), students were required to write a thorough description of how the biological organism or system achieves its function including as much detail as is scientifically known to describe the structural, process, or systems-level workings. To rank as a high-quality ADP (*Y*), students were required to write a broadly translatable, engineering/architecturally oriented, and scientifically accurate design principle that was devoid of biological terminology.

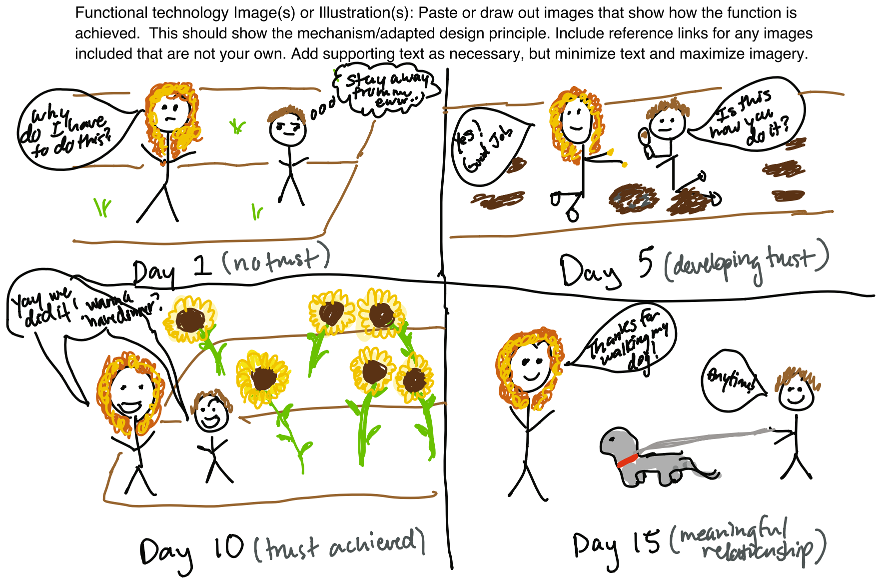


Fig. 8 (continued)

6.3 Phase 3: Statistical Analysis

We utilized 2 × 2 contingency table statistical testing [21] in R to test for (a) independence between the quality of written ADPs and the quality of written mechanisms and (b) independence between quality of written ADP and evidence of hand drawing by students during their NTS process.

6.4 Phase 4: Qualitative Analysis

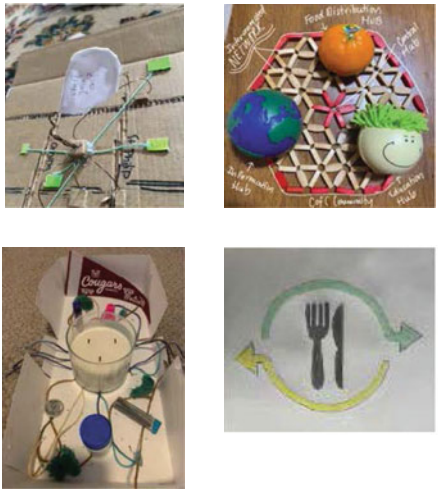
Our weekly roundtable discussions during the co-creation of this manuscript regarding where our students struggled and where they thrived led to qualitative findings and recommendations for optimal biomimicry teaching principles.

7 Results

7.1 Phase 1: Assignment Cues Comparison

We found a large degree of synchrony and overlap in how we facilitated the Nature's Technology Summary (NTS) assignment (Table 2). The background material and learning objectives were alike. Similar exercises and instructional cues

Fig. 9 **a** *Step 6 of NTS,* sweater from honey bee (*Apis mellifera*), A. Feltrinelli, THUAS. **b** Step 6 of NTS, hummingbird *(Trochilidae)*, K. Boakye, ASU. **c** Step 6 of NTS anemone and clownfish M. Gonzales, CofC



Campus food sharing
- Connects pre-existing food resources

Education
- Pre-existing curricular studies
- Food literacy workshops
- Hands-on sustainable agriculture education in on-campus student gardens
- Mentorship program for underclassmen led by upperclassmen

Information
- Connecting people to outside resources
- Online resource
- Space for individuals or organizations to share what they have or what they need

Fig. 9 (continued)

			Rubric											
S.No	Organism	Function	References	Latin Name	Context	Function	Mechanism	BDP / Context illustration	ADP text	Functional Tech illustration ADP diagram (no biology or = no)		Life's Principles	Description	Form, Process or System analogy used
			3+ ref (must have at least 3 references)	correct, (cap, italics but still a 'Yes"if it is in there (?)	complete + biome (characteristics where it exists in Nature)	BM taxonomy (from the sheet diagram)	Thorough description of how the organism achieves its function	organism (correct species) performing the function in context. citation + web address	No biology, engineering/architec turally oriented, well done.	diagram from literature with citation/web address	Hand Drawn	well correlated and matched, three or more LP's included	explains the context, function, strategy, and mechanism, includes in-text citations	
			Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	F / P / S

Fig. 10 NTS rubric

✓				Rubric												
	S.No		Organism	Function	References	Latin Name	Context	Function	Mechanism	BDP / Context illustration	ADP text	Functional Tech illustration ADP diagram (no biology or = no)		Life's Principles	Description	Form, Process or System analogy used
					3+ ref (must have at least 3 references)	correct, (cap, italics but still a 'Yes"if it is in there (?)	complete + biome (characteristics where it exists in Nature)	BM taxonomy (from the sheet diagram)	Thorough description of how the organism achieves its function	organism (correct species) performing the function in context. citation + web address	No biology, engineering/architec turally oriented, well done.	diagram from literature with citation/web address	Hand Drawn	well correlated and matched, three or more LP's included	explains the context, function, strategy, and mechanism, includes in-text citations	
					Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	Y / N	F / P / S
Numbering continued from 2019_THUAS_DwN																
_15	19_15_NTS_		American Beaver		N	N	Y	Y	Y	Y	Y	N	Y	Y	N [1]	F
_16	19_16_NTS_		Butterfly		N	Y	N [2]	Y	Y	N [3]	Y	N	Y	Y	N [4]	F
_17	19_17_NTS_		Carpenter Bee		N	Y	N [5]	Y [6]	Y	Y	Y	N	Y	Y	N [7]	S [8]
_18	19_18_NTS_		Chameleon		N	Y	Y	Y	Y	N [9]	N [10]	N	N [11]	Y	N [12]	P
_19	19_19_NTS_		Darkling Beetle		N	Y	Y	Y	Y	Y	N	N	N	Y	N [13]	F
_20	19_20_NTS_		Emperior Penguin		Y	Y	Y	Y	Y	Y	N	N	N	Y	N [14]	S
_21	19_21_NTS_		Ivy (English)		Y	Y	Y	Y	Y [15]	Y	Y	N	N	Y	N [16]	P
_22	19_22_NTS_		Fig		N	Y	N [17]	Y [18]	Y	Y [19]	Y	N	Y	N	N [20]	S
_23	19_23_NTS_		Hermit Crab		Y	Y	Y	Y [21]	Y	N [22]	Y	N	Y	Y	N [23]	S
_24	19_24_NTS_		Leafcutter ant		Y	Y	Y	Y [24]	Y	Y	Y	N	Y	Y	Y	S
_25	19_25_NTS_		Mangrove		Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N [25]	F
_26	19_26_NTS_		Nautilus		Y	Y	Y	Y	Y	N [26]	Y	N	Y [27]	Y	N [28]	P
_27	19_27_NTS_		Oilbird		N	Y	Y	Y	Y	Y [29]	Y	N	Y [30]	Y	N [31]	P
_28	19_28_NTS_		Oyster		Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N [32]	S
_29	19_29_NTS_		Oyster		N	Y	Y	Y [33]	Y	N [34]	N	N [35]	N	Y	N [36]	P
_30	19_30_NTS_		Prochilodus mariae		N	Y	Y	Y	Y	Y	Y	N	Y [37]	Y	N [38]	P
_31	19_31_NTS_		Pufferfish		Y	Y	Y	Y	Y	N [39]	Y [40]	N	Y	Y	N [41]	P
_32	19_32_NTS_		Rainfrog		Y	Y	Y	Y [42]	Y	N [43]	Y	N	Y	Y	N [44]	P
_33	19_33_NTS_		Reindeer		N	Y	Y	Y	Y	Y	Y	N	Y	N	N [45]	F
_34	19_34_NTS_		Saharan Silver Ant		N	Y	Y	Y	Y	N [46]	Y	N [47]	N	Y	N [48]	F
_35	19_35_NTS_		Saharan Silver Ant		N	Y	Y	Y	Y	Y	Y	N	Y	Y	N [49]	F
_36	19_36_NTS_		Salp		Y	Y	N [50]	Y	Y	N [51]	N [52]	N	Y	Y	N [53]	P
_37	19_37_NTS_		Spiny Orb Spider		Y	Y	N [54]	Y	Y	Y	N	N	N	Y	N [55]	P
_38	19_38_NTS_		Sponge		Y	Y	Y	Y [56]	Y	N [57]	Y	N	Y	Y	N [58]	S
_39	19_39_NTS_		Sunflower		Y	Y	N [59]	Y	N	N [60]	N	N	N	Y	N [61]	F
_40	19_40_NTS_		Sunflower		Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N [62]	F
_41	19_41_NTS_		Termite mound		Y	Y	Y	Y [63]	Y	N [64]	Y	N	Y	Y	N [65]	S
_42	19_42_NTS_		Thorny Lizard		N	Y	Y	Y	Y	Y	N	N	N	Y	N [66]	F
_43	19_43_NTS_		Tree bark		N	Y	N [67]	Y	Y	Y	N [68]	N	Y	Y	N [69]	P
_44	19_44_NTS_		Treebugs		N	Y	N [70]	Y [71]	N [72]	Y	Y	N	Y	N	N [73]	S
Numbering continued in **2019_CofC FYSE**																
					30											

Fig. 11 Sample NTS rubric scoring of one cohort (by external assessor)

Table 2 Synchronizing NTS cues

NTS template elements		
THUAS	ASU	C of C
BS IDE	*BS MS*	*BS BA*
Laura Stevens	*Michelle Fehler*	*Deborah Bidwell*
Organism (common and Latin name)	x	x
Function	x	x
Primary references (minimum 1)	x (minimum 3)	x (minimum 3)
Context	x	x
Biological strategy & mechanism	x	x
Diagram of biological design principle	x	x
(hand drawn)	x	*optional*
Abstracted Design Principle text	x	x
ADP diagram	x	x
(hand drawn)	x	*optional*
	Life's principles	x

x: same as THUAS, BS, MS, and BA (Bachelor of Science, Master of Science, Bachelor of Arts)

were given. The bio-brainstorm research procedures were assigned in an indistinguishable manner, as were the instructions and learning objectives for completing a NTS. All students were assigned the same NTS template listing the organism/ecosystem, the abiotic context in which the biological organism/system evolved, the biological function(s) it is capable of, references from primary resources, a description and diagram of how the organism carries out the biological function, and a description and diagram of the design/engineering abstracted design principle. The slight nuances found in the NTS assignments between institutions were the omission of life's principles at THUAS and the omission of mandatory hand drawn work at CofC.

7.2 Phase 2: Rubric Analysis Results of NTSs

After each NTS was scored, the percentage of students successful for each element of the NTS could be determined (Table 3). Within most of the assignments from the three universities, there were small, overlooked aspects which were common, such as not including in-text citations and not having three or more references (though in most cases, this is not specified within the template). This appears to have had no impact on the overall quality of the NTSs or ADP within the NTS.

Table 3 Included NTS elements

NTS template elements included correctly (*Y*) in assignment (% included successfully)									
THUAS BS IDE	15 week 2019 DwN (%)	15 week 2020 DwN (%)	ASU MS	16 week 2019 DSC (%)	*5 week 2020 DSC (%)	CofC BS BA	14 week 2019 FYSE (%)	14 week 2019 Upper (%)	14 week 2020 Upper (%)
Organism	88	100		97	*100*		100	92	93
Function	81	86		100	*92*		100	97	96
References	50	49		50	*100*		79	65	63
Context	69	50		73	*83*		74	83	67
Mechanism	100	89		93	*100*		100	88	88
Diagram of BDP	94	69		60	*100*		65	77	83
(hand drawn)	94	69		60	*100*		N/A	N/A	N/A
ADP text	94	63		70	*100*		21	40	57
ADP diagram	94	66		70	*92*		11	21	28
(hand drawn)	94	63		70	*92*		5	12	18
Life's Principles	0	0		90	*77*		37	53	52
**Systems	1	14		30	*0*	**	11	21	42

x: hand drawn was optional at CofC
*Short 5-week semester
**THUAS emphasized systems and CofC and THUAS emphasized SDG's in the challenges in 2020

Table 4 Summary of statistical results testing association between quality ADP (*Y*) scores, hand drawing ADP (*Y*), and quality mechanism scores (*Y*) by school and cohort

Results of Fisher's exact tests of association between quality ADP text (*Y*) with quality mechanism (*Y*) and hand drawing ADP (*Y*) pooled by university cohort with CofC FYSE representing control group. Significant results, $a = 0.05$ p values in bold. Odds ratios in parentheses				
	THUAS pooled $n = 51$ NTS	ASU pooled $n = 42$ NTS	CofC upper pooled $n = 201$ NTS	CofC FYSE $n = 19$ NTS
Hand drawn (*Y*)	*p = 0.0007284*	*p = 0.0001367*	*p = 0.000005278*	$p = 1$
	(10.76884)	*(29.80795)*	*(8.804148)*	(0)
Mechanism (*Y*)	$p = 1$	$p = 0.3868$	$p = 0.05253$	$p = 1$
	(0.8740843)	(3.2829709)	(2.618477)	(0)

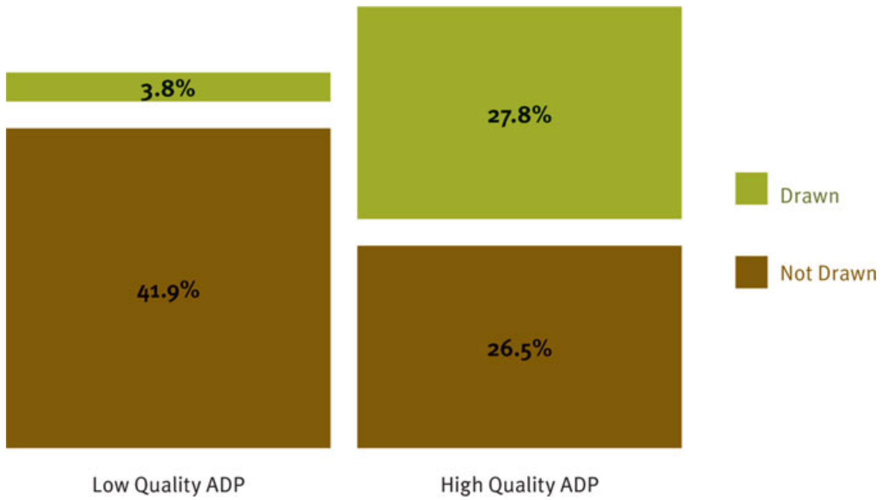


Fig. 12 Summary plot of pooled THUAS, ASU, and CofC upper-class men data showing significant correlation between hand drawing of ADP diagrams and quality of ADP text. Fisher's exact test $p < 2.2 \times 10^{-16}$ Odds ratio = 11.35. n = 313 NTS

Table 5 Summary of statistical results testing association between quality ADP (Y) scores, hand drawing of ADP (Y), and quality mechanism scores (Y) pooled THUAS, ASU, CofC upper level

Results of Fisher's exact tests of association between quality ADP text (*Y*) with quality mechanism (*Y*) and hand drawing ADP (*Y*) pooled THUAS, ASU, CofC upper. Significant results $a = 0.05$ p values in bold. Odds ratios in parentheses	
	Pooled THUAS, ASU, CofC upper. n = 313 NTS
Hand drawn (Y)	*$p < 2.2 \times 10^{-16}$*
	(11.35)
Mechanism (Y)	*$p = 0.0357$*
	(2.343932)

7.3 Phase 3: Statistical Results

Results indicate that CofC first-year experience students performed differently than all other student cohorts and struggled specifically with the ADP step (Table 4). Fisher's exact tests showed no significant differences between upper-level students in 2019 and 2020 cohorts within any university ($p > 0.05$), so data were pooled within schools. Fisher's exact tests show a significant correlation between hand drawing of ADPs and the quality of the ADP text ($p < 0.001$) for all university groups excluding CofC FYSE (Table 4). If there is not a good ADP, it is likely that there was no hand drawing (Fig. 12). There is a trend toward quality mechanisms being associated with quality ADPs in the larger CofC upper-level cohort ($p = 0.05253$), (Table 4). Pooling data across universities (excluding CofC FYSE)

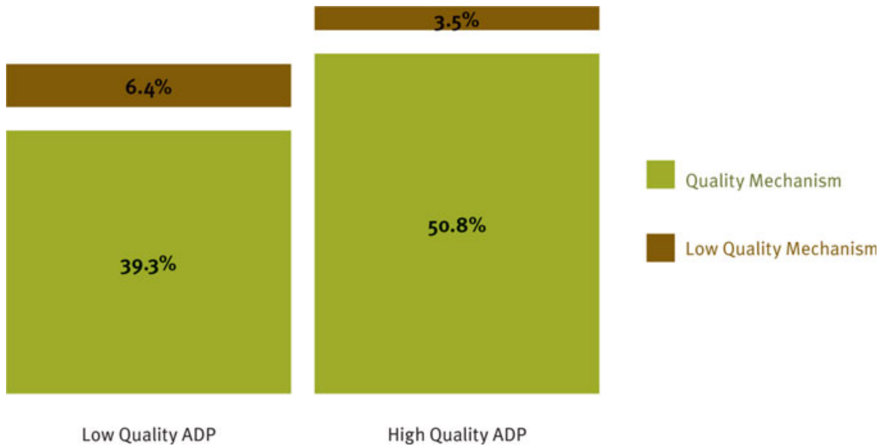


Fig. 13 Summary plot of pooled THUAS, ASU, and CofC upper-class men data showing significant correlation between quality of mechanism and quality of ADP text. Fisher's exact test $p = 0.0357$ Odds ratio = 2.34. $n = 313$ NTS

and applying Fisher's exact testing resulted in a highly significant correlation between quality of ADP text and hand drawing of ADP $p < 2.2 \times 10^{-16}$ (Table 5). Pooling across universities (excluding CofC FYSE) also revealed a significant correlation between quality of mechanism and quality of ADP text ($p = 0.0357$), (Table 5). If there was a quality mechanism, a quality ADP text was more likely (Fig. 13).

7.4 Phase 4: Qualitative Analysis Results

In the final phase of analysis, we gathered elements influential in developing high-quality biomimicry design principles by testing those elements leading to high-quality NTSs (Table 6). Key insights from the three previous phases (left column) were then tested in a single cohort at ASU (right column) with the proposed improvements described below.

Influential course/NTS elements: With THUAS students, none of the NTS submissions included life's principles, though this is due to these not being included as a section, as was the case with non-mandatory in-text citations. Although the assignment was not mandatory and not assessed as such, more than half of the students delivered. Here, the 'Context' section lacked, and many opted for background information only, obviating the need to describe the habitat and environmental conditions these champions face daily. Students choose to draw the biological organism more often than the abstracted principles, frequently jumping to the design phase before this knowledge was internalized. Charleston students' contexts were generally clear and complete, providing an understanding of the

Table 6 Findings table elements influential in high-quality BMY design principles

Elements influential in achieving high-quality NTS’			
THUAS and CofC 2019–2020, ASU 2019			ASU 2020 Test group +
Key insights comparison Phase 1	– Life’s principles THUAS – No required three references THUAS – Hand drawing CofC – Not all students did same # NTS’	Influence additions	LPs were added to help understand the deeper patterns in nature
Key insights from NTS assignment results similarities/differences Good/bad Phase 2	– ADP texts speculative and vague CofC (partially all) – Some ADP diagrams include biological elements (which they should not) – Jumping to design solutions – Content missing THUAS – NTS sections often missing THUAS where biological drawings explaining how mechanism works (jumping to design) – Speculative desired characteristic mechanics – CofC FYSE students struggled with ADPs	Proposed improvements	– ADP iterations in class – Partial NTS assignments, iterative, short feedback loops prior to next step – Underlying theme of function, strategy, mechanism throughout – Assign biomechanical drawing as separate assignment for emphasis – Peer feedback to introduce repeating exposure to observations in nature and drawn translations – Require the drawing of biology by hand, rather than finding an illustration – Narrow focus of NTS to one organism, process or system
Key insights systems Phase 2	– Fewer systems in 2019 (not requested) – THUAS and CofC stressed attempting systems in 2020 class	Proposed improvements from phase 1 course elements	– Progress from form, to process, to systems as course proceeds – Iterations system mapping of challenge – NTS examples of systems relationships

(continued)

Table 6 (continued)

Elements influential in achieving high-quality NTS'			
THUAS and CofC 2019–2020, ASU 2019			ASU 2020 Test group +
			– Explain text and hand draw BDP diagrams with exactness from scientific research – Future proposal to start with form, transgressing to process and then system analogies
Key insights Phase 3	– Direct correlation between high % hand drawing and achieving 'good' ADPs – Direct correlation between good mechanism and achieving 'good' ADPs – Correlation high % correct organism, function, context, etc., to achieving systems when 'organism' chosen is a system	Proposed improvements	– *Multiple* ADPs/NTSs – Require hand drawing – Require feedback on mechanism before ideation
Main suggestions for improving ADP in NTS in biomimicry methodology – Develop NTS in rounds: (1) learn biology, (2) draw BDP, (3) ADP text (feedback to refine), (4) draw ADP diagram, and (5) all fields added – Setup timeline for multiple short feedback loops – Choose drawing assignments studying form, process, and systems to mimic, fitting design function – Select only 1 strategy from 1 organism, behavior, system function per NTS – Hand draw BDPs showing exact mechanism referred to from literature – Possibly move from emulation of form to process to systems as course progresses – Add/keep LP's and min. three references			

conditions; however, they were not required to hand draw the biology or abstractions themselves and often had biological explanations where an abstraction was required.

All students demonstrated a struggle with both the abstracted design principle texts as well as the diagrams depicting these. ADP texts in many cases were vague and either referenced the biological mechanism or the champion, failing to give

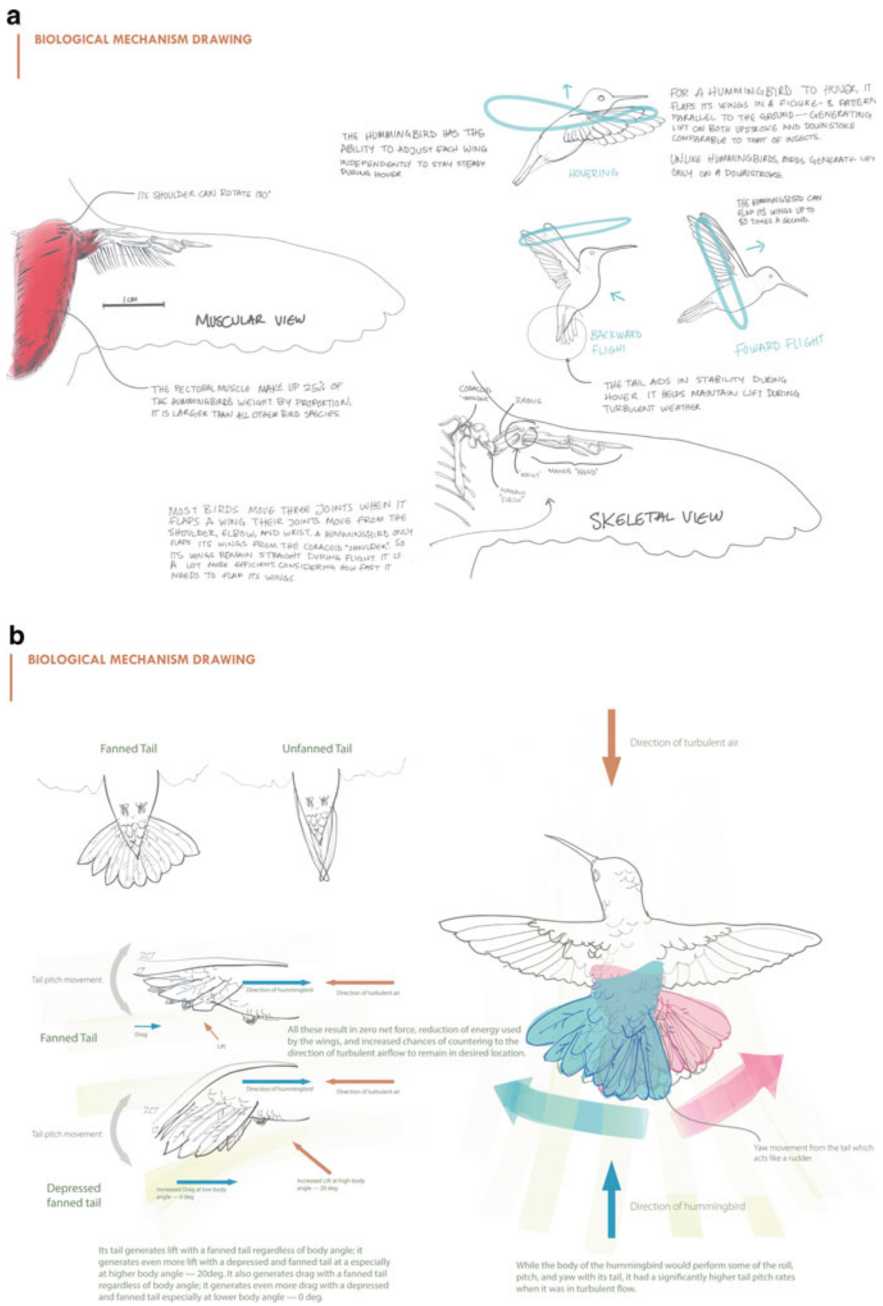


Fig. 14 **a–c** Example of the evolution of the biological mechanism drawing for the hummingbird NTS biomechanical drawing process between feedback loops. K. Boakye, ASU summer 2020. Continued feedback and research allowed the student to iterate, refine, and deepen his understanding of the mechanism. This allowed him to arrive at a stronger ADP

c

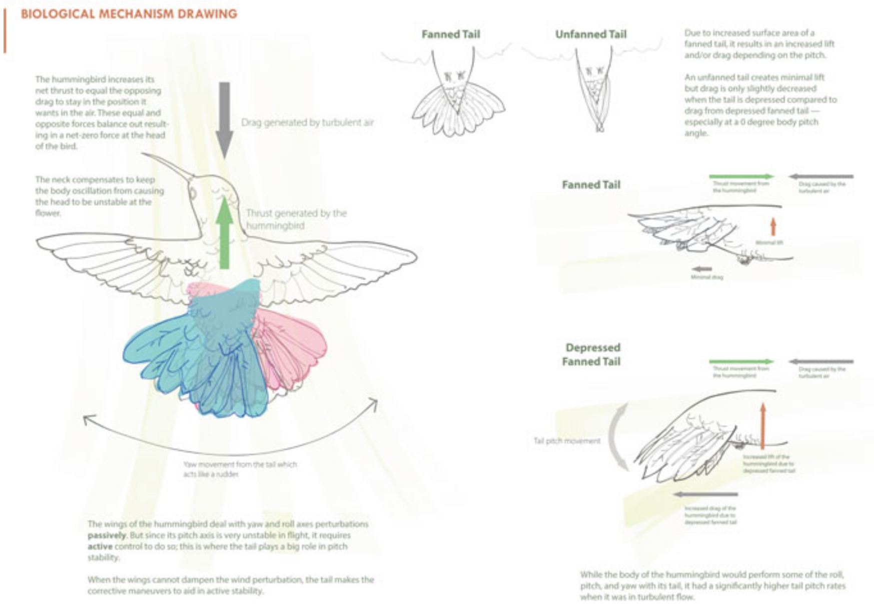


Fig. 14 (continued)

insight into how the mechanism functions, while others jumped ahead to begin thinking of design solutions within the ADP rather than the more open-ended statements which would aid in the development of multiple design ideas. CofC students tended to speculate in the ADP on the desired characteristics that the champion displayed for use in their solution, or simply mentioned that an aspect should be considered. ADPs of both THUAS and ASU students in design-oriented semesters seemed to do quite well, while fewer Charleston students had success. In many cases, CofC students added a depiction of the biological champion achieving the mechanism rather than something 'abstracted' from biology or nature to the ADP. CofC students had the majority of systems-level ADPs ready to start the ideation/emulating phase of design thinking (Table 3).

Note that the ASU 2020 cohort had the benefit of what was being learned during this study and was added last to test our 'recommended pedagogical principles' at an early stage. Our analysis shows that the highest overall scores were achieved for this cohort (except for the number of system analogies which were not requested by the instructor) (Table 3). The ASU summer cohort, being a short 5-week class, was not exposed to the full breadth of the BMY thinking process. A biology to design approach was chosen, where the students picked an organism of choice based on fascination or curiosity, which then inspired their NTS focus. Evidence for effective learning through multiple iterations of biological mechanism drawings with frequent instructor and expert feedback can be seen in Fig. 14.

(a) First biomechanical drawing focused on the wings of the hummingbird. At this phase, the student continued to learn about the mechanism of the stabilization of the bird.
(b) Once the student discovered that the tail was a more important strategy in stabilizing, he drew a new biological mechanism drawing
(c) After more research, feedback, and two interviews with experts, the student updated the drawing to be more specific about how the mechanism worked. This deeper understanding led to a stronger ADP.

8 Discussion

This study examines characteristics, methods, factors, descriptors, learning outcomes, and techniques which are most often present in biomimicry student work that correlates to the highest-quality abstracted design principles for the creating phase of biomimicry design thinking.

8.1 RQ1

In this study, we looked at what subcomponents of the biomimicry thinking methodology are most vital for students to achieve high-quality abstracted design principles. We found that each element of the NTS template appears to be valuable. Identifying natural history, context, and function aligns with the scoping phase. Strategy and mechanism are vital to translating nature to design during discovery and key to accurate emulation during brainstorming. Including life's principles and requiring references from the scientific literature may help students accurately describe the biological mechanisms, and accurate biological mechanisms are correlated with stronger ADPs (Table 3, Table 5, and Fig. 13).

First-year experience (FYSE) students can be thought of as a control group, a novice population. First semester US college students are in transition from high school to university-level academics and typically lack the foundational knowledge and skills for excelling at the same higher-order thinking level as their upper-class men counterparts [22, 23]. The NTS assignment for the FYSE cohort emphasized achieving a quality mechanism through an introduction to primary literature research more so than emphasizing achieving a high-quality ADP. Although taught how to generate ADPs in an identical manner as CofC upper-class men, the FYSE cohort's ADPs were not adequate. Although the FYSE cohort excelled at achieving quality mechanisms (Table 3), they were largely unsuccessful at ADP writing. Understanding the biology alone appears inadequate for translating it into design principles. While our best learning comes when we are stretched and challenged, quality long-term learning requires frequent low stakes opportunities that provide retrieval practice, repetition, and interleaving [24]. Our results support the growing body of pedagogical knowledge suggesting that students need iterative practice and frequent feedback when developing new skills such as defining biological functions using the biomimicry taxonomy, writing a succinct biological strategy, or researching and interpreting the biological mechanism underlying a form, process, or system in nature [25, 26]. The influence of class size on pedagogy is an important consideration. In larger classes, utilizing guided peer review in lieu of more time-consuming instructor-only feedback may be a valuable tool for providing rapid, iterative assessment of early NTS steps in both online and face-to-face courses [27].

It is worth reminding our readers that during the spring semester of 2020, we experienced a global pandemic caused by a novel coronavirus. College campuses around the globe abruptly shifted from face to face to online learning during this COVID-19 outbreak. The ASU biomimicry summer 2020 class was online by design, but THUAS and CofC courses in spring 2020 experienced a disruptive shift from face to face to online learning, in the middle of the team-based challenge to biology design project.

During this study, we could apply what we discovered from our early findings to integrate research and action, adapting our pedagogy in real time [28]. Challenges identified in facilitating the NTS template in 2019 and early 2020 courses were addressed in the ASU summer 2020 cohort test group. We found that presenting the NTS assignment in four steps with multiple (peer) feedback moments on each was linked with improving the abstracted design principle quality to 100% (Table 3).

8.2 RQ2

We also wanted to discover what curricular or pedagogical factors influence whether students achieve systems-level abstracted design principles. THUAS, ASU, and CofC emphasized challenges that encouraged students to attempt system analogies, while the five-week semester at ASU could not do this in 2020 because of time constraints. It is possible that a higher percentage of systems-level ADPs was achieved by the 2020 CofC class because the challenges centered on UN Sustainable Development Goals (SDGs), which are wicked problems and triggered more systems-level approaches to ideation. On the other hand, THUAS also focused on challenges concerning the SDGs in both years but had a lower percentage of students who were successful in systems thinking. We noted that several CofC students included mutualistic, symbiotic relationships within their NTS submissions. Encouraging students to seek out organisms that exhibit mutualism is likely to be a good pedagogical strategy for achieving systems-level NTSs in future cohorts. Systems-level biomimetic emulation will be vital for becoming a regenerative species. ‘The first step is to imagine it, to envision this symbiotic world, a world in which we are a welcome species—a nature contributor’ [29]. We are interested in exploring the possible pedagogical benefits of beginning students with function, strategy, and mechanism explorations-based first on forms, then expanding to processes and then to systems in progression over the course of the semester.

8.3 RQ3

We finally asked: What is the art behind the science of biomimicry? Does drawing to learn improve the learning outcomes of multidisciplinary novice biomimicry practitioners? If so, how is it correlated with achieving high-quality abstracted design principles? The biomimicry thinking process has been explained in detail in the Biomimicry Resource Handbook [4] as well as in biomimicry step by step [5]. They both mention sketching or hand drawing as part of iSites, which are guided nature journaling observations. Sketching is usually taught as part of a biomimic's practice [17], and it could be implied that sketching does help with understanding nature's strategies. However, hand drawing had not previously been specifically highlighted as part of the process to arrive at a high-quality ADP. Our study shows that it should be. We discovered that the use of hand drawing to depict the visuals in the NTS is highly correlated with high-quality abstracted design principles.

The vital step of hand drawing the biological mechanism is a discovery we feel is a contribution to biomimicry pedagogy and was not entirely expected. Drawing helps to peak curiosity while allowing the biological information to be internalized by the learner [13]. Emphasizing the intended purpose of drawing as an inquisitive learning tool and process can help students overcome their fear of drawing. 'Artistry is not a prerequisite for most uses of drawing as a tool' [13]. We also want to highlight the importance of repeated low stakes practice with prompt feedback and opportunities for iteration while discovering and sketching the biological mechanism (Fig. 14).

The correlation between a quality mechanism and a quality ADP, though weaker, should not be overlooked. We note that many students could achieve quality ADPs without including hand drawings in their final NTS files. These students likely did draw to learn the mechanism (which was emphasized by the instructor) but did not feel the quality of their drawings merited inclusion in the final draft of the CofC NTS. It should also be considered that students have diverse learning strategies, and while drawing to learn may work for the majority, accommodating other ways of knowing and learning is advisable. Reaching a high-level understanding of the biological mechanism is the target, and drawing to learn is correlated with high-level achievement of this learning outcome. However, our recommendations are not prescriptive best practices, but flexible and adaptive suggestions for effective pedagogical principles.

To better emphasize the biological mechanism drawing, it became its own assignment (instead of it being part of the NTS) for summer 2020 ASU students who were required to submit their hand drawn assignments as letter-sized PDF documents to increase the focus on understanding of detail, scale, and perspective. Low stakes peer feedback on visuals allowed rapid proof of understanding if the drawing clearly explained the function, strategy, and mechanism with a sufficient level of detail, and what improvements were needed. Additional feedback was provided by the instructor and students revised their drawings for NTS part 3. The biological mechanism drawing clearly assisted students in their learning process (Fig. 14), but also turned out to be one of the most important steps for the instructor to gauge the level of students' understanding of the organism and is highly recommended as an optimal pedagogical principle. Given a compressed schedule, the B2D process served as an effective way for learning how to abstract design principles for the first time. Future research could explore the difference in learning the biomimicry thinking process through a comparison of students who research form-based, process-based, and systems-based biological mechanisms. This may help determine whether instructors should intentionally move students along this paradigm from form to process to systems in their pedagogy.

Providing iterative feedback on the components of the NTS has been shown to be beneficial. Our results indicate it would be wise to add an additional step (step 5) between writing the ADP and drawing the ADP. We believe it to be important that the ADP is written well first, then the drawing can become a test of the writing. We recommend students team up in small groups to share their written ADP drafts with their peers who then make a first attempt to draw them. This low stakes interactive process was tested with a team of four CofC spring 2020 students during an online check in and shown to help identify any gaps that are lost in translation and allow for rapid ADP text prototyping based on peer feedback. When a written ADP draft passes the peer 'draw-it' test, it is likely well written enough to be useful to designers and engineers during the creative innovation phase. After iterating the written ADP with peers, students can then draw better ADP diagrams. More testing is needed to determine if these higher-quality ADPs will translate into improved design outcomes.

9 Concluding Remarks

Our research aimed to discover which characteristics, methods, factors, descriptors, learning outcomes, and/or techniques are most often present in biomimicry student work and correlate to the highest-quality design principles. We have come to understand the greater importance of the Nature Technology Summary as an essential tool for biomimicry designers and the crucial aspect of dividing it up in sections to be handled separately. In attempts to follow the biomimicry thinking design cycle completely, some students may have rushed through this assignment. It has become clear, this may not be rushed, and the NTS needs a more prominent place in the curriculum before the design phase can begin. While we have seen students struggle with moving from this exercise to the start of the design phase before [3], taking the time and getting feedback in several rounds have been a major discovery. We have also learned that just asking students to focus on systems analogies between biology and design helps, but asking specifically for symbiotic relationships helps this far more. We think that requiring students to focus on forms, process, and systems in a consecutive order may help. Each comes with different challenges. Forms are easier to see, but require a lot of detail to understand the mechanics. Processes may require deep dives into biochemistry which can be overwhelming for students lacking STEM backgrounds. The system is very complex to understand, but the design principles are metaphorical which require less detail/resolution of mechanisms. We can conclude that hand drawing both the biology and the abstracted design principles triggers a deeper connection to the researched organism or system and will continue to test these pedagogical principles in the future.

- Biomimicry students struggle with the translating phase of biological mechanisms to design principles, which are necessary for biomimicry design solutions.
- To help students overcome obstacles of this translation, the Nature's Technology Summary exercise is most optimal when divided in sections with intermediate feedback sessions.
- Hand drawing improves the translation phase and helps students internalize the science.
- Consecutively addressing forms, processes, and system analogies in biomimicry design may help students understand the differences between each, but more testing is needed in this field.

Acknowledgements We would like to acknowledge the ease at which this team came together to work across time zones and expertise. Our passion for biomimicry was the driver, and our empathy for each other is the connecting thread. We are also deeply humbled by the genius of all organisms around us, the limitless inspirations they shower us with and the awe we are left with whenever we open our eyes and truly see. 'Perhaps in the end, it will not be a change in technology that will bring us to the biomimetic future, but a change of heart, a humbling that allows us to be attentive to nature's lessons' [29]. A special thanks to our students, without whom, this research would not have been possible.

Compliance With Ethical Standards The authors declare no conflict of interests. The Hague University of Applied Sciences and Delft University of Technology—Netherlands Code of Conduct for Research Integrity (2018). College of Charleston IRB-2020-06-04-183355—Protocol Exemption Integrated Science Volume 1 Chapter Art and Science of Biomimicry. Arizona State University HRB 503A Social Behavior Protocol Integrated Sciences vol. 1.

References

1. Benyus JM (1997) Biomimicry—innovation inspired by nature. HarperCollins, New York, NY USA
2. Biomimicry38 (2020) Biomimicry 3.8
3. Stevens L, Kopnina H, Mulder K, DeVries M (2020) Analogies in biomimicry. Int Technol Stud
4. Baumeister D (2014) Biomimicry resource handbook. Biomimicry 3.8
5. Rowland R (2017) Biomimicry step-by-step. Bioinspir Biomim Nan. https://doi.org/10.1680/jbibn.16.00019
6. Harris P, Ambrose G (2009) Basics design 08: design thinking. Bloomsbury Publishing, Switzerland
7. Casakin H, Goldschmidt G (1999) Expertise and the use of visual analogy: implications for design education. Des Stud 20(2). https://doi.org/10.1016/S0142-694X(98)00032-5
8. Landin J (2015) Rediscovering the forgotten benefits of drawing

9. Ainsworth S, Prain V, Tytler R (2011) Drawing to learn in science. Science 333(6046):1096–1097. https://doi.org/10.1126/science.1204153
10. Dietrich A, Kanso R (2010) A review of EEG, ERP, and neuroimaging studies of creativity and insight. Psychol Bull 136(5):822–848. https://doi.org/10.1037/a0019749
11. Fan JE (2015) Drawing to learn: how producing graphical representations enhances scientific thinking. Transl Issues Psychol S 1(2):170–181. https://doi.org/10.1037/tps0000037
12. Dove JE, Everett LA, Preece PF (1999) Exploring a hydrological concept through children's drawings. Int J Sci Educ 21(5):485–497. https://doi.org/10.1080/095006999290534
13. Quillin K, Thomas S (2015) Drawing-to-learn: a framework for using drawings to promote model-based reasoning in biology. CBE Life Sci Educ 14(1). https://doi.org/10.1187/cbe.14-08-0128
14. Van Meter P, Aleksic M, Schwartz A, Garner J (2006) Learner-generated drawing as a strategy for learning from content area text. Contemp Educ Psychol 31(2):142–166. https://doi.org/10.1016/j.cedpsych.2005.04.001
15. Zhang HZ, Linn M (2010) Proceedings of the 9th international conference of the learning sciences. Paper presented at the international conference of the learning sciences, Chicago, 29 June to 2 July 2010
16. Kang H, Thompson J, Windschitl M (2014) Creating opportunities for students to show what they know: the role of scaffolding in assessment tasks. Sci Educ 98(4):674–704
17. Rovalo E (2019) iSites nature journaling for biomimicry. Biomimicry 3.8, Missoula, Montana
18. Pyle C (2000) Art as science: scientific illustration, 1490–1670 in drawing, woodcut and copper plate. Endeavour 24(2):69–75
19. Leslie CW, Roth CE (2003) Keeping a nature journal: discover a whole new way of seeing the world around you. Storey Publishing, LLC, United States
20. Khakpour A (2012) Methodology of comparative studies in education. Contemp Educ Res J 1:20–26
21. Fisher RA (1934) Statistical methods for research workers. Oliver & Boyd, Edinburgh
22. Chaffin L, Korey CA, Warnick C, Wood L (2019) Life so far: a grounded theory analysis of college student transitions through the sophomore year. J First-Year Exp Stud Transit 31 (2):79–93
23. Bloom BS (1956) Taxonomy of educational objectives: the classification of educational goals. Longmans, Green, New York, NY
24. Brown P, Roediger III HL, McDaniel MA (2014) Make it stick. Harvard University Press
25. Roediger HL III, Pyc MA (2012) Inexpensive techniques to improve education: applying cognitive psychology to enhance educational practice. J Appl Res Mem Cogn 1(4):242–248
26. Biwer F, Oude Egbrink MGA, Aalten P, De Bruin ABH (2020) Fostering effective learning strategies in higher education—a mixed-methods study. J Appl Res Mem Cogn 9(2):186–203. https://doi.org/10.1016/j.jarmac.2020.03.004
27. Allen D, Tanner K (2005) Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. Cell Biol Educ 4(4):262–268
28. Salafsky N, Margoluis R, Redford KH, Robinson JG (2002) Improving the practice of conservation: a conceptual framework and research agenda for conservation science. Conserv Biol 16(6):1469–1479
29. Benyus JM (2020) Biomimicry Institute Fireside Chat

Laura Stevens holds two MS degrees in the fields of Architecture from Delft University of Technology and Biomimicry from Arizona State University. She is a biomimicry design educator in her role as a professor in the Industrial Design Engineering program at The Hague University of Applied Sciences in the Netherlands. A sustainable design instructor since 2007, she is presently working on her Ph.D. in Delft, where she is writing a series of peer-reviewed articles and chapters on the topic of biomimicry design thinking as a methodology to enhance circular, systems thinking solutions in design by learning from time-tested biological strategies and mechanisms found in nature. *Many thanks to the co-authors, Michelle Fehler, Deborah Bidwell, and Asha Singhal, all of whom dedicated their knowledge and experience as professionals in the field. The chapter is an example of beneficial mutualism and would have been impossible without them.*

The Phenomenon of Meeting in Art: A Transdisciplinary Approach

30

Natalya Shelkovaya

… высвобождаясь от власти малого, беспамятного «я», увидишь ты, что все явленья – знаки, по которым ты вспоминаешь самого себя («… when you become free from the power of small and unconscious "I", you will see that all the phenomena are signs that help you to remember yourself». M. Voloshin "Apprentice".).

М. Волошин "Подмастерье".

Summary

The chapter is devoted to revealing the deep meaning of the meetings, which fill our life, but being locked in the shell of our ego and our problems (often invented by ourselves); we ignore what is often vital for us. Using the examples of paintings by brilliant artists (Leonardo da Vinci, Raphael, Jusepe de Ribera, Karl Bryullov, Alexander Ivanov), the author analyzes the phenomenon of meeting through the personal insights of the author, who at a certain stage in her life, saw and felt what she had not seen and had not felt before. A meeting occurs when the souls of those who meet come in touch. In art, this means to feel what the Creator of the masterpiece felt. However, often people look without seeing, listen without hearing, and therefore do not understand the meaning of the signs sent to them through various meetings with people, works of art, etc., which are often "codes" of a person's fate came to realize his or her essence and mission on

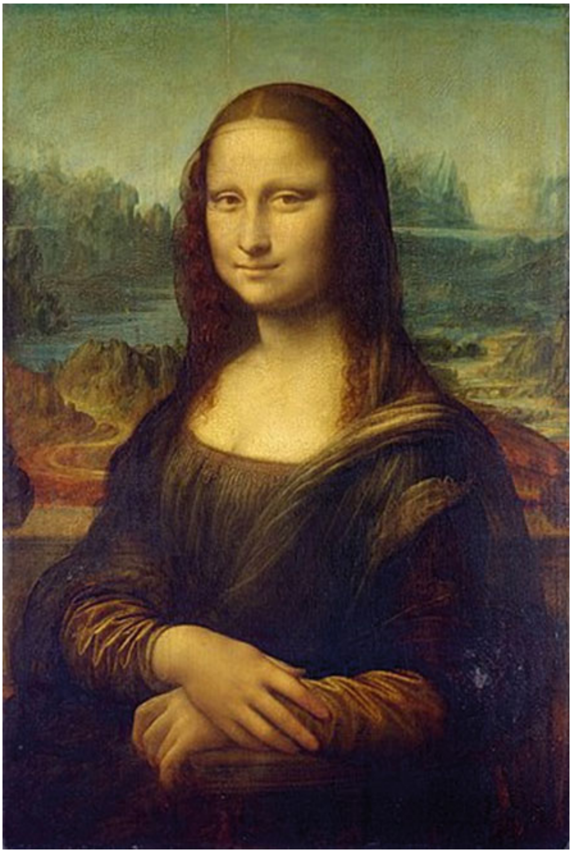
N. Shelkovaya (✉)
Prospekt Tsentralnyi 59 a, Severodonetsk 93406, Ukraine
e-mail: shelkovaya@snu.edu.ua

Volodymyr Dahl East Ukrainian National University, Severodonetsk 93406, Ukraine

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Severodonetsk, Ukraine

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_30

earth. If the meeting took place, then it can transform a person's consciousness, worldview, and feeling of the world. The author reveals such a transformation by analyzing her meeting with Mary Magdalene in the painting by A. Ivanov, *The Appearance of Christ to Mary Magdalene*, concluding the holiness of human corporeality and the holiness of human love, not darkened by the fall into selfishness and animality.



Leonardo da Vinci. Mona Lisa. Who is depicted in this picture: Lisa del Giocondo, the image of Leonardo himself as an integral androgynous personality, or his state of fullness from love for this woman? Where does the magnetism of this portrait reside? Maybe, it is because when we look at her, we do not see a picture, but a real woman with fabulous female charm?

(*Adapted from Wikimedia Commons, the free media repository* https://en.wikipedia.org/wiki/Mona_Lisa).

The code of this chapter is 01001101 01100101 01100101 01110100 01101001 01101110 01100111.

1 Introduction

What is a meeting? A meeting is, first of all, an intersection of the senses of those who meet. It is an attraction of the like to the like and spiritual, heartfelt contact of the points of likeness. A meeting is a sign, an indication, a hint, for each of those who meet, at something relevant (sometimes very important). However, are we able to see, hear, understand, and finally, guess by intuition the meanings of these signs? Alas, we are often not.

Meetings are different: pleasant and unpleasant, desired and undesired, memorable (sometimes for the whole life) and disappearing from our memory immediately after leaving. But, there exists *special* meetings: meetings that drastically change our worldview and world outlook and/or “imprint” as a symbolic sign into our destiny. To my mind, such kinds of meetings can be regarded as the meetings with the masterpieces of art, or rather, with the souls of their creators that somehow touch our souls and often unconsciously become “codes” of *our* destinies. This touch (emanating from the artist) is sometimes slight, or gentle; sometimes, it leads (i.e., guides) to the most profound shock of the whole human nature, turning one's life upside down (or maybe just *returning to oneself*, after being lost in the mundane bustle?).

2 Meeting the Mona Lisa

A meeting with a masterpiece of art… What is a masterpiece? What are the criteria? Why do we meet with some masterpieces of art, while others pass by our souls?

To answer these questions, I will resort to my own cases of experiencing meetings with the souls of artists. For a definite time, I have been meeting not only people but also books and paintings that hint at something very important for me. Such meetings pervade my whole life.

Since I am a woman, most of the paintings I met and the souls I touched with were depicting women. The first woman of this kind was La Gioconda (Graphical Abstract).

The meeting with the Mona Lisa by Leonardo da Vinci was preceded not only by a spiritual and profound joyful expectation of seeing a masterpiece of art but also by an ordeal of standing in line for 12 hours at the Pushkin State Museum of Fine Arts in Moscow. When I entered the hall, I hoisted two pairs of glasses on my nose to see well. However, I did not have to examine it… As I approached the picture, Mona Lisa's gaze was more and more perceptible, and against my will, I plunged deeper and deeper into the realm of female charm emanating from her smile. I was walking past the picture, or rather, past a *real* charming (i.e., enchanting, captivating everyone's look) woman, and she turned her face toward me with a little coyness, looked at me… and smiled. Oh, that was a true woman! Moreover, her beauty was not in the external beauty of the face but in the conquering female charm. I was stunned…

For a long time, my impression of this picture remained in my soul as a meeting with a *real* woman who gave me a marvelous (from the word “marvel,” miracle) smile. While preparing this article's text, I gathered the information about the “Mona Lisa,” which is much discussed nowadays. I mean that Leonardo da Vinci painted himself; his spiritual state of fullness of being, self-sufficiency, and I had the following idea. Leonardo did not paint himself but his *beloved woman*. “Bathing” in the rays of her smile and melting in her female charm, he experienced such bliss… This state of bliss close to nirvana, a state of fullness of being (as in Goethe’s “Faust:” “*Beautiful moment, do not pass away!*” [1, p. 52] or Blake’s phrase: “*Hold Infinity in the palm of your hand*”… [2, p. 490.]), which arose from the energetic Yin-Yang fusion (the combination of the energies of love between a woman and a man), was reflected by a genius in his masterpiece. The “Mona Lisa” is a song of love that bestows a state of harmony and fullness of being.

My *meeting* with the Mona Lisa eventually transformed into *communication* with her, continuous communication. At a certain stage of my life, a challenging stage, I wanted to reach a peaceful state of mind and self-sufficiency. And I, not knowing why (but can we know *all* the reasons for our actions?), began to keep the image of Mona Lisa in my mind's eye to keep it all the time, scrutinize it, or rather, the state that she emits, deeper and deeper. Gradually, I became calmer and entered “the Mona Lisa state,” i.e., the state of peace, self-sufficiency, and fullness of being. Later, I learned that Buddhist monks act similarly, *contemplating* the image of one or another divine being in order to acquire the qualities that it possesses (Not observe, “look at …,” but contemplate, thereby, according to V. Dahl, “*смотреть со смыслом, вникая, углубляясь в предмет <...> вникать во что мысленно, разумом, духом*”[1] [3, p. 261]). However, I *was ignorant* about this type of psychotechnology when I chose it unconsciously (by intuition). Maybe, the Mona Lisa helped me?

3 Meeting the Holy Inessa

Yet, the meeting with the Mona Lisa was for me, first of all, getting a meeting with the *true* woman and the *true* person, possessing a state of self-sufficiency, the fullness of being, and, as a result, peace (note: the face of a truly loving and loved radiates not only the Light of happiness but also profound peace of mind and serenity), I touched the *genuine* holiness while I was in the Dresden Gallery. It is traditionally believed that the pearl of this gallery is the “Sistine Madonna” by

[1] “look attentively with meaning, delving into, deepening into the subject, to delve into mentally, examining, admiring”.

Fig. 1 Raphael. Sistine Madonna. Who can we see in this picture: Madonna and Christ child or a very young woman who anxiously presses her child to her chest and does not want (!) to give Him in the world to be torn to pieces by the crowd? *Adapted from* https://en.wikipedia.org/wiki/Sistine_Madonna

Raphael, and I was looking forward to this very meeting, particularly with this picture. But, life had other plans…

The Sistine Madonna (Fig. 1) was beautiful, perfect, but she was not divine at all. I was astonished by the childish fear on the face of the Virgin Mary, who clutched the baby to herself and looked reluctant to give Him to the world. It was not the exalted heavenly Mother of God, but an earthly, very young mother-woman. Nevertheless, my experience of contemplation of the "Sistine Madonna" did not touch me to the core. The miracle of the meeting did not happen. I went to another hall slightly upset and suddenly…

Suddenly, my look was captivated, as if with a magnet, to the painting of Jusepe de Ribera "*The Holy Inessa*" (Fig. 2). I could not take my eyes off this picture, repeating to myself: "*Poor, wrongfully convicted creature!*".

Let me recall the plot of the picture. In III century, a young Roman girl adopted Christianity, unbeknownst to her parents and refused to worship pagan gods. In punishment for this, she was slandered, accused of perversity, and put naked in the central square of the city. God saw the wrongfully convicted suffering and performed a miracle: Inessa's hair grew up quickly to her toes, covering her naked

Fig. 2 Jusepe de Ribera. The Holy Inessa. Why was a young and innocent girl, still a child, put in the central square of the city naked as a punishment? Why did her hair grow up immediately to her heels, and an angel brought her a white (symbol of purity) cover? Why does the crowd often punish the purest and most spiritual people? *Adapted from Wikimedia Commons, the free media repository* https://commons.wikimedia.org/wiki/File:Jos%C3%A9_de_Ribera_027.jpg

body, and an angel from heaven covered her with a white cloth. What eyes St. Inessa had! I will never forget this look! Purity, heavenly purity, and pain, the inhuman mental anguish of the wrongfully convicted… Many years later, I experienced something similar and realized why it was *this* picture that attracted me…

4 Meeting with Karl Bryullov's Feeling of Love

It was another stunning meeting. More precisely, I had three meetings in the Tretyakov Gallery. I did not use to love K. Bryullov for the polished beauty of his paintings. For example, "*The Last Day of Pompeii*" (Fig. 3) is a terrible human tragedy. Meanwhile, we admire a beautiful (dead!..) woman in the foreground, in the center of the picture… But, this time I *saw* something else (often we look, but we do not see, we listen, but we do not hear).

I stood next to Bryullov's "*Daughters of Pacini*" (in Russian, the painting is known as "*The Horsewoman*") (Fig. 4), glancing over the stunning beauty of a woman and a little girl, and suddenly, I *felt* the moods that Bryullov experienced

Fig. 3 Karl Bryullov. The Last Day of Pompeii. The terrible eruption of Vesuvius. The beautiful city of Pompeya is dying. People are terrified. In the foreground, there is a beautiful young dead woman with a little child crying near her. Yet, what do we feel when we look at the painting and this woman? Is it horror? Or is it admiration for the beauty of this woman? If it is admiration, why? What is the difference between beauty and prettiness, in particular, in paintings? *Adapted from Wikimedia Commons, the free media repository* https://en.wikipedia.org/wiki/The_Last_Day_of_Pompeii

when he painted this picture. I *felt* his love for this woman. How much he loved her! (You probably *know* about it. I *did not know* about his love, but I *felt* it). Turning my eyes to the right, I *felt* again Bryullov's enormous love for this woman, depicted in another picture, "*Italian Midday*" (Fig. 5).

Bryullov did not portray beautiful women, but *his love* for one woman, Countess Yuliya Samoylova! Why haven't I felt this before? Maybe, because, I was not able to feel the energy of love?

The energy of love works wonders and turns a work of art into a masterpiece. I had this idea not so long ago when I saw again the paintings of K. Bryullov, now in the Russian Museum in St. Petersburg. Even from far, I paid attention to his picture "*Crucifixion of Jesus*" (Fig. 6).

When I still saw the picture vaguely, and its details were indistinct for me, I felt the suffering and pain emanating from the picture. Coming closer, I saw and felt that this suffering and pain did not come from Christ but from Mary Magdalene, whose prototype was, of course, the artist's beloved Countess Yuliya Samoylova. I stood in front of the picture and could not take my eyes off Mary Magdalene. Yes,

Fig. 4 Karl Bryullov. Daughters of Pacini, Giovannina, and Amazilia. Who is depicted by Bryullov in the image of a beautiful horsewoman: The daughter of Pacini or the countess Yuliya Samoylova who he loved immensely? If it is Pacini's daughter, then why does the energy of Bryullov's love for Yuliya Samoylova radiates from this painting's contemplation? *Adapted from Wikimedia Commons, the free media repository* https://commons.wikimedia.org/wiki/File:1832._BRULLOV_VSADNICA1.jpg

it was not Christ, but Mary Magdalene, the main character in this picture. What emotional eyes she had! Amid tears, her eyes expressed immeasurable suffering and were looking into eternity where Christ has gone. I stood for a long time peering into the eyes of Magdalene, going away from the here-and-now-being and plunging into eternity together with her. Why didn't I pay attention to this picture before? Maybe because of being immersed in here-and-now, I could not notice the "look into eternity" to glance at the eternity with Magdalene?

Crossing the hall with other paintings of K. Bryullov and peering into examining the characters' eyes on the paintings, I found that the genuine masterpieces were only the paintings that represented the beloved woman of K. Bryullov, Countess Julia Samoilova. Moving from one hall to another, I felt again and again that *only great love turns a skillfully performed work of art into a masterpiece*. This cannot always be love for a woman. It can also be love for the sea, as with Ivan Aivazovsky, and love for the forest, as with Ivan Shishkin, and love for the night, as with Arkhip Kuindzhi, and love for sunlight, as with Auguste Renoir.

Fig. 5 Karl Bryullov. Italian Midday. Why, looking at a beautiful Italian woman with a bunch of grapes, in the rays of the sun that caresses and admires her, the image of Countess Yuliya Samoylova appears in our memory? *Adapted from Wikimedia Commons, the free media repository* https://en.wikipedia.org/wiki/Karl_Bryullov

5 Meeting with the Soul of Alexander Ivanov

Let us return to the day, I first met K. Bryullov's Art at the Tretyakov Gallery in Moscow. Leaving the "Bryullov hall" for the "Ivanov hall," I met with the soul of another artist, Alexander Ivanov, on the paintings "*Christ's Appearance to Mary Magdalene after the Resurrection*" (as fate would have it, the painting then was exhibited in the Tretyakov Gallery, while usually, it is in the Russian Museum in Saint Petersburg) and "*The Appearance of Christ Before the People*."

I was looking at the "*Appearance of Christ Before the People*" (Fig. 7) and heard Ivanov's soul silent cry: "*People! What have you done! How could you crucify Him? How can you crucify Him again and again, spreading hatred, enmity and evil instead of love and forgiveness?*" It was a cry from the heart of a profoundly believing, *true* Christian who *truly* loved Christ. Alas, people did not hear his cry… People continue to crucify Christ, not knowing how to love, forgive, and care for the other, *thinking only about themselves and forgetting about the other*, staying locked inside the shell of their "Ego."

Hardly had I recovered from the meeting with such overwhelming faith, I turned around and saw Christ and Magdalene and encountered the wonder.

Fig. 6 Karl Bryullov. Crucifixion of Jesus. Who is the main character in the picture? Who emanates immense suffering? Christ? No, it is Mary Magdalene. Her eyes. Her look into eternity, where Christ has gone. These eyes and the image of countess Yuliya Samoylova reappears. Why? *Adapted from Wikipedia, the free encyclopedia* https://upload.wikimedia.org/wikipedia/commons/b/b3/%D0%91%D1%80%D1%8E%D0%BB%D0%BB%D0%BE%D0%B2_%D0%A0%D0%B0%D1%81%D0%BF%D1%8F%D1%82%D0%B8%D0%B5.jpg

6 Meeting the Wonder

At the entrance to the Tretyakov Gallery, on the main staircase, I felt so excited by the beauty that filled me with light and joy, the feeling of meeting with the wonder. Yes! a true wonder. For the nature of wonder lies first of all in its secret, mystery, inscrutability.

The modern Western rationalized civilization has sterilized; according to C. Jung, everything is mysterious, secret, irrational, transcendent, belonging to the sphere of the unconscious, with its mind. At the same time, it impoverished the human psyche; having rejected from the consciousness, everything that cannot be understood, considered, reasoned, thus turning a person from the great and unfathomable mystery into an object and subject of cognition, a programmed

Fig. 7 Alexander Ivanov. The Appearance of Christ Before the People. Why do the faces of people who saw Christ not express joy but surprise and fear? Are they *not ready to meet* Him? Why? Are we *ready to meet* with Him? *Adapted from Wikimedia Commons, the free media repository* https://en.wikipedia.org/wiki/The_Appearance_of_Christ_Before_the_People

biorobot. Its human has lost the ability to see wonder and fantasy in everything. Entering the Tretyakov Gallery, I felt that I had come to a fairy tale.

What is a fairy tale? It is nothing more than the reality represented with the help of symbols. The one who knows how to decode symbols *meets* with the heroes of a fairy tale and *draws* wisdom, which enters not only consciousness, but also fills the sub- and supra-conscious sphere. This is something similar to the character of knowledge that Socrates spoke about: when it comes inside a person, it *wholly transforms them,* and they no longer *know* about justice, kindness, wisdom, but *become* themselves just, kind, wise. The perception of a fairy tale, as well as a myth, requires *entering* the fairy tale world, the syncretism of thinking, the temporary loss of the "I" in the "other world." Little children can do that, ancient people could do that, but we are often *unable* to do that. Why? Our awareness of the "I," "ego" interferes and creates a barrier between us and the fairy tale or myth ("This is fiction!"), between a person and the world. Meanwhile, we do not notice that we do not obtain but lose ourselves and the world, living in the world, and inside the "I" we have invented.

Why does the oriental wisdom say: "*Be as wise as a serpent and as pure as a child*"? Not just pure, but as a child. As a child's worldview is unaffected and lively, it is always ready to see a miracle everywhere and in everything, and therefore, it perceives the world with the *eyes wide open* (and ears too). Such a perception is able to *see* and *hear*, not just look and listen; hence, it is wise. Not smart but wise. Looking back through the ages, we do not admire the brain of ancient people, first of all, but their wisdom, which is often beyond your understanding.

Mind and genius can be evil and cruel; their center is the head, while wisdom is the unity of the heart and the head. Moreover, priority is given not to the head, brain, but the heart. The ancient Egyptians reckoned that "*heart is the seat of intelligence.*" Jesus Christ always paid attention to this in his preaching: "*You will indeed hear but never understand, and you will indeed see but never perceive… For this people's heart has grown dull, and with their ears they can barely hear, and their eyes they have closed, lest they should see with their eyes and hear with their ears and understand with their heart*" (Matthew 13: 13, 15).[2]

Are we able to, and do we want to *see* the wonder in everything, to see the God, enchanted in the world, as R. Steiner put it [4, p. 24]? Can a biorobot be surprised and *believe* in miracles? If someone *loves* nature, he/she is united with the world. Then, nature and the world give their *love back* and unveil their depths, their laws. Thus, appeared the "wonders" of the ancient Egyptians, Mayans, Aztecs, etc.

If someone *blinded* by the pride of the "I," "ego," tries to *dominate* the world (force it); then, the world answers with the same: it "closes" their eyes and ears, and they lose the ability to *produce and create* in *harmony* with nature and lose the ability for a wonder.

Nevertheless, the world we live in is the greatest *wonder*.

7 Christ and Magdalene: Holiness of Corporality

The one who *touches* the wonder *meets* the world and the *signs* that it sends to us. One such sign was my *meeting* with Ivanov's painting "*Christ's Appearance to Mary Magdalene after the Resurrection*" (Fig. 8).

What was so special about this picture that attracted me and then surprised me? It was Magdalene. Christ was wonderful (in terms of the artist's skillful performance) but usual (there was no wonder in him). But Magdalene, she was beautiful, sensual, appealing, and holy. How is it possible to describe *this*: sensuality and holiness at the same time? And that not the hypocrite holiness of the sanctimonious, but genuine holiness, purity, spirituality. Can we imagine a sensual icon? Hardly. However, Ivanov's image of Magdalene is what it is, a sensual icon of Magdalene, to pray to her and to desire her, then to shudder from the physical desire for the saint. How *could* Ivanov portray this? How could he render a *wonder*?

[2] Hereinafter, the Bible is cited according to the version of its translation on the Bible Gateway website; URL: https://www.biblegateway.com/.

Fig. 8 Alexander Ivanov. Christ's Appearance to Mary Magdalene After the Resurrection. After Ivanov created this masterpiece, his friends advised him not to paint anymore because it was impossible to paint something better, for Ivanov depicted something unimaginable—the miracle of the unity of corporality, physical desire for a woman, and the holiness, the reverent awe in front of her in the image of Mary Magdalene. But, is this unity a miracle or a consequence of our fall? Could Christ be in love with Magdalene as a woman with holistic spiritual, mental, and physical love? Can love for Magdalene `stain' Christ? Is love a sin? Is not love given to people by God? *Adapted from Wikimedia Commons, the free media repository* https://commons.wikimedia.org/wiki/File:Alexander_Ivanov_-_Christ%27s_Appearance_to_Mary_Magdalene_after_the_Resurrection_-_Google_Art_Project.jpg

Perhaps, the unity of holiness and beautiful physicality is a miracle only for us sinners, for whom the body is of the devil, the source of temptation, sin, and vice; for whom, since the fall of Eve and Adam, man has been *split* into body and soul (spirit), while the body is *drawn* to earth (sin), and spirit is drawn to Heaven (God). Was it always like this? Is this the real nature of human corporality? Why did Christ communicate with prostitutes, and Mary Magdalene became His *favorite* disciple? Moreover, according to the Apocrypha, she was also his beloved woman. Thus, the apocryphal Egyptian Gospel of Philip (Coptic text) says: "*The [Lord loved] Mariam more than [all the (other)] Disciples, [and he] kissed her often on her [mouth.]*" (Philip 59) [5].

The body itself is neither sinful nor blessed. Those are the features of the human spirit. If a person's spirit is deprived (that is, the spirit is rotten, "dead," is smoldering, decaying), then the body is deprived, perverted, sinful.

A person with a holy spirit, living *for* God and *with* God, has a holy body. In this case, they "*do what they want*" (Augustine) and *remain* holy, for they *cannot* do or conceive of anything contrary to the will of the Creator. At this point, faith in God is not the remote relationship "God–me," but the organic unity of God and "I," the constant feeling of God in the spiritual heart. The vigilant "guardian" of this unity, the voice of God, is a conscience that constantly "talks" with a person (conscience is silent in fallen people).

In the case of the falling away (fallen person) from God, it is the spirit that falls away first of all, and as a result, there follows the reservoir of the spirit and soul—the body—as it is only a "vessel" of the spirit and soul. The paradox of the unnatural (from the point of view of the Creator, creation) correlation of spirit and body is that the spirit turns from the leader, teacher, keeper of the body into its slave. The body, which is *not given from Above the knowledge* of the correct way of human development (for it is the mission of the human spirit), "rushes" to the abyss of being human and brings there the spirit…

Moreover, the body of a person deprived of spirituality becomes bestial, worse than an animal, which we see in the modern, so-called civilized Western world (while the east has also become Western, as paradoxical as this may sound). However, the human spirit without flesh is more an angel than a human. Only, the unity of spirit and flesh constitutes a genuine person, a person with heart, soul, and flesh, as a unity of the earthly body and heavenly (Divine) spirit. This must be exactly what the Christian Saint Athanasius the Great said: *"For the Son of God became man so that we might become God"* [6]. Christ came into the world to show that the *Divine Spirit* can live in a *human body*. Man is the unity of the earthly and divine.

"*And the Lord God formed man of the dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul*" (Genesis 2: 7). Who breathed? God did. Who gave a soul? God did. Where does this "divine soul" live? In a human body.

In my opinion, this is what Christ wanted to show, and this is what A. Ivanov *could* feel and convey in his painting "*The Appearance of Christ to Mary Magdalene*."

The traditional ideal of a Christian is the ascetic monk, who voluntarily renounces all worldly blessings, bodily pleasures, and first of all, marriage. Why? After all, the first blessing of God to Adam and Eve to all people was as follows: "*Be fruitful and increase in number*" (Genesis 1: 28). So, people who deviate from the first commandment blessing of God, therefore, from God, believe that they will thus be closer to Him? This is a paradox, is it not? Are marriage and love sin and not godly? Can love be a sin if God is Love? Either "God is love" (1 John 4: 8) is not true, or we do not understand love in the right way, except the fact that we do *not know how to love…*

Yes! The deviation of Eve and Adam from God, who is love, was a deviation (since we have deviated, we have been deviating, deviating, deviating further and further, further, and further…) from love. Moreover, Christ was sent to *return* people to God, i.e., to love, *to teach people to love*. Christ taught us to love by his example. He communicated with everyone: with people of his faith and foreigners, men and women, tax collectors and sinners, and even with prostitutes. This provoked surprise and misunderstanding, and "*when the Pharisees saw this, they asked his disciples, Why does your teacher eat with tax collectors and sinners?* On hearing this, Jesus said, *It is not the healthy who need a doctor, but the sick* (Matthew 9: 11–12). Christ was the healer of human souls. He cleaned their "dirt" with his purity, without fear of "getting dirty" for himself. They approached him as sinners but left him as saints.

In connection with this, I would like to recall a Taoist parable. Two monks traveled together and saw a young and beautiful woman who stood in front of a stormy stream, afraid to cross it. One of the monks picked her up in silence and carried her over the stream. Then, both monks walked their way for a long time in silence. Finally, the second monk asked: "*How could you carry that woman when we took a vow as monks not to touch women?*" The first monk looked at him, smiled a wise smile, and replied: "*I set her down and left her there, but you are still carrying her*."

What is the essence of asceticism? Renunciation from the flesh, wasting of the flesh. It means that the ascetic is consumed with his flesh. An ascetic monk, who despises his body, considers it the leading cause of all human sins. He fights with his body, torments, and kills his flesh to "make room" for the spirit and get rid of the temptations of the flesh. Spirit is exalted; the flesh is tamed. However, scouring the flesh and the body, a man shows that he does not love himself. Moreover, since God gave spirit and soul and body, God *knows best what* to give to man. Moreover, rebellion against the body is, therefore, rebellion against God.

In this case, there is an illusion of freedom from the body, an illusion of freedom of the spirit, since such a person is completely absorbed in the problem of killing his flesh, torturing the body. Such a person is completely body-centered, body-oriented, and therefore dependent on the body.

They say that the great Christian Saint Francis of Assisi loved everyone, treated people, animals, birds, and fish as "brothers," but at the same time exhausted his body with asceticism. Then, his followers, Franciscan monks, told him: "*Brother Francis, you love everyone, but you do not love your body and torment it.*" He replied, referring to his body: "*My brother body, forgive me, I will no longer torment you and I will take care of you.*" Since then, Francis ceased to waste his flesh, driving it almost to death.

One of the main commandments of Christ is: "*Love thy neighbor as yourself*" (Matthew 22: 39). That is one must *learn* to love oneself in order to love others.

A true saint is one who radiates the divine light of the soul and illuminates with it everything, including his body, for the body is the house for the soul and spirit, and the house is where you are loved. The asceticism and wasting of the flesh are matched more to Platonism, where the body, by courtesy of Plato, is regarded as the prison for the soul, than to Christianity. Christ *never* was an ascetic. *Having fasted for forty days in the desert,* having prepared himself spiritually for the teacher's role, he no longer fasted, which repeatedly raised complaints from Jewish priests. In response to reproaches about non-observance of fasting, Christ replied: *"What goes into someone's mouth does not defile them, but what comes out of their mouth that is what defiles them"* (Matthew 15: 11).

Christ never was an ascetic: he ate and drank wine, talked to all people, did not despise even harlots. Moreover, one of them, Mary Magdalene, became his most faithful, devoted, and good disciple. "*The other [women] saw his loving Mariam, they say to him: Why do thou love [her] more than all of us? The Savior replied, he says to them: Why do I not love you as (I do) her?*" (Philip 59) [5].

Why did Christ love Mary Magdalene more than anyone? She was the only one in His lifetime who *learned to love*. Love lived in her heart even before meeting Christ, but it was perverse physical love, not the spiritualized love of the flesh. Having met Christ, not only her spirit and soul but also her flesh was transformed. Moreover, she became His follower that is in the literal sense, followed His footsteps.

Once, she learned that Christ came to one of the noble houses, and he was not even greeted according to the customs of that time (feet washing). She, as poor as she was, bought expensive myrrh, washed His feet with her tears, wiped them with her hair, anointed them with myrrh, and kissed them, constantly repenting. The owners were indignant at how Christ allowed her, the sinner, to be in the noble house and touch Him. Crist said: *"Therefore, I tell you, her many sins have been forgiven—as her great love has shown. But whoever has been forgiven little loves little"* (Luke 7: 47). She did love sinfully, but she loved (!!!), and the true divine love could enter her heart. Moreover, for those who have *never loved anyone, how can they love God?*

A friend in need is a friend indeed. What did the disciples (friends) of Christ do during His most terrible ordeals and torments, and what did Mary Magdalene do in this situation?

Christ, who was not an ascetic, once again, his last farewell evening at the Last Supper, eats and drinks wine. He predicts the betrayal of Judah. What did the disciples do to prevent betrayal? Nothing. Moreover, when Christ asked some of them after the Last Supper to pray with him in the Garden of Gethsemane, they fell asleep. He woke them up, asked them to support him (to support Christ!..), and they were falling asleep again and again. Will you fall asleep if the one you love asks you to support him at the most difficult moment of life? However, they fell asleep…

Further on, Judas kissed Christ with a kiss of betrayal (there are so many Judas' ``kisses of love" in the world!..), and Jesus was caught, judged, scourged, and crucified. Were His disciples near? Did they even support him with a look? Peter, his favorite disciple, was denounced three times because of fear; the rest disappeared somewhere (where?).

Whom was with Christ during the crucifixion? Two Marries—Mary Magdalene and Mary, the mother of James and Joseph. Whom was the first to come to visit the body of Christ after his crucifixion? Mary Magdalene. Whom did the angel tell about the resurrection of Christ? Mary Magdalene. Who immediately believed it? Mary Magdalene. Did the disciples immediately believe (recall Thomas's disbelief)? Who did the angel instruct to carry the message of the resurrection of Christ? Mary Magdalene. The angel *knows better who to* entrust to carry *such* a message.

Could *such* a woman be unworthy of the love of Christ? She was worthy. And He loved her and "*[and he] kissed her often on her [mouth]*" (Philip 55) [5]. In "Pistis Sophia" ("The Wisdom of Faith" (Greek)—the name of a Coptic manuscript from Egypt), a beautiful, loving sinner is given more than forgiveness: "*Mary (Magdalene), blessed one, the fullness, or all-blessed fullness, thou who shalt be sung of as blessed in all generations*" [7]. But, this information is on the Apocrypha and is not recognized by the Church. You would object; it is not recognized by the Church indeed. However, does this mean that it does not correspond to reality? In any case, the authenticity of the Apocrypha is not denied.

Indeed, can divine love be a sin? Is it, once again, a sin to love a woman with both body and soul? Why is physical love a sin? The animal copulation of man and woman for pleasure without love is lust, sin, terrible sin; definitely, this is a kind of necrophilia. True. However, is genuine, spiritual, heartfelt, and physical love a sin? Moreover, why do we associate love between a man and a woman purely with the body, sex? Perhaps, the reason for this is our split nature, inferiority, and lack of spirituality, inability to love *like people*, or rather, like God—integrally, with spirit, heart, and body?

So why did Christ come into this world? To be a *sacrifice* to God for human sins or a *teacher*? The teacher of divine love, lost by Eve and Adam due to their fall, lapse from virtue and God–love that means lapse from true love, divine, integral, spiritual, and physical love altogether. One of the consequences of this sin is a person's loss of wholeness, splitting into body and soul, body and spirit, loss of spirituality by the body, which leads to the animalization of a person (which we see with the naked eye today).

In the Gospels, Christ is considered the second Adam. Thus, in the genealogy of Jesus (Luke 3: 23–28), Christ is a direct descendant of Adam. The Epistle of Paul to the Romans speaks of Christ as the second Adam and "*a pattern of the one to come*" (Romans 5:14). Is the "pattern of the one to come" according to God's plan, the "crucifixion" of humanity? In the "ideal" chapter of the Bible, i.e., from the book of Genesis, Chapter Introduction to Integrated Science: Transdisciplinarity, which tells about the ideal, from the point of view of God, structure, and way of life of the world and man, it is said that God created man and woman and blessed them: "*be fruitful and increase in number*" (Genesis 1: 28), i.e., a man was not created for an ascetic lifestyle, and a woman was not created for a monastery.

So, what is the mission of the Messiah? The most common version is the vision of a sacrifice to God in Christ, the lamb (sacrificial lamb) of God. This view of the mission of the Messiah is at the cornerstone of the Christian Church. One of the main dogmas of the Christian symbol of faith is the dogma of redemption; according to which, Christ came to this world in order to atone for human sins with his sufferings and torments on the cross, first of all, to atone for the original sin of Eve and Adam. The Church, in particular, relies on the Gospel of John, which states: *"For God so loved the world that he gave his one and only Son, that whoever believes in him shall not perish but have eternal life"* (John 3: 16). This reminds me of the cruel Phoenician religion, where children were sacrificed to the God Moloch. Moreover, the whole concept of Christ as a sacrifice is very similar to pagan sacrifices. In general, what ‟nice' love it is: if you torture My Son to death, I will forgive your sins… This implies the cult of the crucifix, the cross's cult, the cult of suffering, and the cult of death Christians honor the great martyrs…

Is this the only and true version of the mission of Christ? Who was Christ: The sacrifice or the Spiritual Teacher? The Church does not deny that He is a Teacher. However, why did this Teacher come? Atone for human sins? But how? What is the essence of sin, including the original? Sin is a fall from God, a fall from divine love. The *meaning of original sin* is in egoism (and the Church does not deny this), for Eve placed *her* desire above God's desire, and standing under the tree of the knowledge of good and evil, she thought only *of herself*, of blessings *for herself*: *"When the woman saw that the fruit of the tree was good for food and pleasing to the eye, and also desirable for gaining wisdom…"* (Genesis 3:6). An egoist thinks only of himself but does not know how to love and does not even love himself because *thinking about oneself and loving oneself is not the same thing. Then Christ came to rid us of selfishness and teach us not to "talk about," "think about" ourselves and about love, but to love.*

"Rejoice and be glad" invokes Jesus in His Sermon on the Mount (Matthew 5: 12). Rejoice at everything and always be thankful to God for everything, be *great rejoicers*! Love God and every creation of God. Such love is stronger than death; it conquers death. Love is a cult of life, a cult of resurrection, the victory of life over death, the spirit over the body, the eternal over the temporary, and the momentary. Do you remember the beautiful words of the apostle Paul about love? *"If I have the gift of prophecy and can fathom all mysteries and all knowledge, and if I have a faith that can move mountains, but do not have love, I am nothing… Love is patient, love is kind. It does not envy, it does not boast, it is not proud. It does not dishonor others, it is not self-seeking, it is not easily angered, it keeps no record of wrongs. Love does not delight in evil but rejoices with the truth. It always protects, always trusts, always hopes, always perseveres"* (1 Corinthians 13: 2, 4–7).

If this is true, then can the love of Christ for Magdalene be sinful? Can corporality be a saint? Can one feel the desire for the saint? And here, the question arises: *"What is holiness?"*.

Traditionally, holiness is associated with asceticism. However, a saint is not an ascetic. The ascetic is "attached" to his body; he is "consumed" with the killing of the flesh (as mentioned above). The saint is free from body dependence. The main thing in holiness is not asceticism but the divine light of the soul (in Russian, the words “light” and “holy” derive from one etymological stem). When this light is kindled in the soul, then it transforms both the soul and the body. This was always paid attention to in hesychasm (the mystical tradition in Orthodoxy), saying that during the act of deification (theosis), there is a combination of the uncreated nature of God and the created nature of man, which leads to a *complete* transformation of man. Deification is the return of man to God, to himself in the original form, to the man living *for* God and *with* God he was *before* the fall.

This peculiarity of holiness is remarkably described by F. Dostoevsky in “The Brothers Karamazov,” using the example of the Saint Elder Zosima, who loved to indulge in tea with jam, chatted with the church members with pleasure, but at the same time was a *true saint*. Meanwhile, there lived a silent monk next to him, who led a visibly ascetic lifestyle. His asceticism was in terms of physical, corporal needs, but the envy corroded his soul for the Elder Zosima, who was loved, while he was not.

Therefore, can a person live in the world, have a family, communicate with people, rejoice with great and small pleasures, and be holy at the same time? It is possible! Moreover, it is the saint-living-in-the-world who is the pattern that Christ brought to us. The Indian guru Bhagwan Shri Rajneesh reflected upon this as follows: “*Christ loved life indeed, like all people, He loved all its colors, He could enjoy life, but at the same time He was a saint, highly spiritual man. He could be close to the “dirt” without getting dirty, be close to the darkness, but spread light, be the Light*” [8, p. 149].

Christ did live in the world, celebrated weddings (recollect His miracle of turning water into wine at the Marriage at Cana), spoke to everyone, even harlots, and was a saint, and he *loved sacredly* one woman, Mary Magdalene, loved her with all his heart, with all his soul, with his mind. Mary answered him with the same entire spiritual, heartfelt, and corporal love, too.

Yet, Christ was crucified; they *crucified the divine love*. Information about His love for Magdalene was put in the Apocrypha. And a woman, just as before the appearance of Christ, continues to be considered the source of evil, the temptress, in general, the one who can lead to sin and not to holiness. Finally, let me repeat once again, the connection, the *Divine Connection* with a woman, the connection, named *divine love*, does not lead to sin (fall from God's Grace), but to integrity, androgyny, holiness (coming to God). Maybe, this is what Christ wanted to show? Maybe, this is why the angel (the messenger of God) instructed the woman, Mary Magdalene, to bring the message of the resurrection of Christ?

Today, in the so-called Western civilized society, in the era of the post (or post-post?)-modernism, when the cult of corporality flourished, as never before, in theory (philosophical and literary reflections on corporality), and in mass media, as well as in life, the *inferiority* of a human body devoid of spirituality is especially felt. The splitting of a person into body and soul is felt as never before, echoing with the tragedy of divine love's crucifixion, its transformation into animal copulation, beautifully named “making love.”

The feeling of the sacred, holy, is more and more lost. The very *ability to love is lost*. Is it possible to talk with animals and biorobots about holiness? Can most modern people understand and feel the *holiness of human corporeality, the holiness of human love?*

I stood in front of A. Ivanov's painting "*The Appearance of Christ to Mary Magdalene*" deeply immersed in these thoughts for a long time. I was interrupted by the guide, who approached the painting with his excursion. I started listening to his story. The guide said that when Ivanov created this masterpiece, his friends were shocked by this painting and advised him not to paint anymore because it was impossible to paint something better. Moreover, he followed their advice for a long time. However, the voice of God may have appealed to him to create "*The Appearance of Christ Before the People*." God gave Ivanov the inspiration for these two paintings so that people would *hear* Christ and get inspired by the faith to *see* Magdalene and remember the unfortunately lost holiness of corporality. However, people did *not* hear and did *not* see… Why? Why have I never felt the most profound faith emanating from the painting "*The Appearance of Christ Before the People*," and did not pay much attention to the painting "*Christ's Appearance to Mary Magdalene after the Resurrection*"? I was not ready for these meetings before.

8 The Essence of the Phenomenon of Meeting

This is where the sacramental question raises: "*Is it possible to be close to a masterpiece of art and not feel it?*" It is. Moreover, many people *do not feel* this. Perhaps, they *know* that they are *next* to the masterpiece, but they *do not feel what* emanates this artwork. They do not feel *something the Creator felt* while crafting his work. Does a *meeting* with a masterpiece happen in this case? No, it does not. One only increases the *amount of knowledge* about this work of art. While this happens in the brain, *head*, the meeting takes place in the soul, *heart*.

In contrast with science, based on *abstractions*, art, and life, *living* people, *living* world, cannot be understood only with the head's help; they can only be *felt*. One can slightly touch their inner nature, i.e., to *themselves*, with a slight touch, gentle, with awe, with reverence to the great mystery of the world, of man, and art.

Why are modern people, especially those belonging to Western civilization, increasingly losing (or have lost?) the ability to feel life, living people, and masterpieces of art, i.e., genuine art? Why are *the meetings of souls and hearts* so rare in today's unlimited information space? An answer to this question was given back in the first century AD by Christ, who said that "*You will indeed hear but never understand, and you will indeed see but never perceive. For this people's heart has grown dull, and with their ears they can barely hear, and their eyes they have closed, lest they should see with their eyes and hear with their ears and understand with their heart and turn, and I would heal them"* (Matthew 13: 13, 15).

It is not the head, but the heart that is the center of the human essence. It is in the heart that deep thoughts and feelings are born ("deep" as they are not "on the surface˵—in the head, but "in-depth"—in the heart). The West's disregard of the "culture of the heart," which is pertinent to the East, leads to the inattention to the "connecting link," which is the heart. It lies between "the top"—the mind, the spiritual core, and "the bottom"—the physiology, the body. There is no spiritual, cordial, and physical harmony, for the heart animates both "top" and "bottom" and constitutes the human. Without this connecting and spiritualizing "link," spirituality transforms into rationality, can evolve up to robotics, and corporality into beastliness, which we can see, alas, in the modern world.

The world opens only to the one who takes a good look at it. Over the past two hundred years, this look has become more and more machine-like and inhuman. Paul Virilio notes in his work "The Vision Machine" that the outlook has been dehumanized. He concludes that modern technocratic civilization replicates the vision without a look, finally, the reproduction of blindness, industrialization of the absence of look [9], i.e., production of biorobots.

Can an animal or a biorobot feel like a masterpiece of art? Can they *meet* it? No, they cannot. Otherwise, can a biorobot or an animal *create* a masterpiece of art? No, because this does not require knowledge and information, but inspiration, spiritualization, which are born in the *heart*. While the heart of the modern Western "civilized"(?) people is covered with an "impenetrable shell…".

9 Heart as a Meeting Place with God, Love, and Art

It is the heart that is the *meeting* place with God, love, and art. When this meeting happens, God and Love unite into a single whole in art naturally.

One of the world's most brilliant dancers, Isadora Duncan (Fig. 9) wrote about the genuine art of dance: "*It is a prayer, this dance; each movement reaches in long undulations to the heavens and becomes a part of the eternal rhythm of the spheres.*" However, "*the dance of the future will have to become again a high religious art as it was with the Greeks. For art which is not religious is not art, is mere merchandise*" [10, p. 173, 175] (the way we, alas, often see it in modern society as an entertainment show). Once, after Isadora Duncan's dance, she was approached by the composer, whose music was brilliantly improvised in her dance. He said: "*Incredible! How did you know how I felt when I wrote this music, how did you manage to convey what I felt with your dance?!*" She replied: "*I do not know. When I listen to music, I "immerse" into it with my whole being and some mechanism (engine) "turns on" at the base of my spine. It controls my body despite my consciousness.*" Despite my consciousness, both the act of creating a work of art and the creative act of its perception occur beyond consciousness in the heart, not in the head. "*I do not know why I wrote like this," have been repeating the true creators of masterpieces for millennia. "I do not know why this particular work of*

Fig. 9 Isadora Duncan as the first fairy in Midsummer night's Dream. Can dance be a form of prayer? Maybe it should be like that considered the greatest dancer in the world, Isadora Duncan. "*It is a prayer, this dance; each movement reaches in long undulations to the heavens and becomes a part of the eternal rhythm of the spheres—as Duncan out it.* <...> But the dance of the future will have to become a high religious art again as it was with the Greeks. For art which is not religious is not art, is mere merchandise." *Adapted from Wikipedia, the free encyclopedia* https://uk.wikipedia.org/wiki/%D0%86%D1%81%D1%96%D0%B4%D0%BE%D1%80%D0%B0_%D0%94%D1%83%D0%BD%D0%BA%D0%B0%D0%BD#/media/%D0%A4%D0%B0%D0%B9%D0%BB:Isadora_Duncan_-_first_fairy.jpg

art overwhelmed me so much. I do not know why I fell in love with this person," people say again and again.

I do not know, I do not know, I do not know… This is the formula of wisdom (let us recall Socrates). The "awareness of ignorance" is an introduction to understanding the great mystery of Man, World, and Art, the great mystery of the Creator, who let masterpieces appear in the world.

10 Conclusion

And yet, if the meeting did occur, it is not only and not so much the "merit" of the heart that can feel the Other, the world, the art, but the "sign of fate" sent from Above. Moreover, the person has to decipher it. Sometimes, this sign is the keynote (at the time of the meeting, it is often unconscious) of a person's life. It happened to me precisely this way. Pictures of Leonardo da Vinci, Jusepe de Ribera, Alexander Ivanov brought out from the layer of my subconscious something that I had been going to from youth, when I wrote in my notebook: “What is Truth?” at the age of 15, without hearing the word “truth” *from anyone else* and thus set myself on the path to truth, God, love, holiness. Later, I have lived many years, suffered, and felt a lot, and I had an idea: *everyone takes their way to truth, God, and love, and when they reach them, they understand that God, Truth, and Love are not three, but one*. The great mission of the masterpieces of art is that they help to come to this…

Core Messages

- The meeting is the “sign of fate” sent from Above.
- The heart is the *meeting* place with God, love, and art.
- The heart of the modern Western “civilized” people is covered with an “impenetrable shell.”
- The masterpieces of art help a person open his heart and meet the vital signs of his destiny.

References

1. Goethe, J.W.: Faust. In: Luke, D. (trans.). Oxford University Press, Oxford (1987)
2. Blake, W., David, V.E. (eds): Erdman. In: The Complete Poetry and Prose of William Blake (Newly revised ed). New York, Anchor Books/Doubleday (1988)
3. Dahl V (1998) Tolkovyy slovar’ zhivogo velikorusskogo yazyka (Explanatory Dictionary of the Modern Great Russian Language: In 4 vols), vol 4. Russkii yazyk, Moscow
4. Steiner, R.: Misterii drevnosti i khristianstvo (The Mysteries of Antiquity and Christianity) (trans: from german). Interbuk, Moscow (1990). German edition: Steiner, R.: Das Christentum als mystische Tatsache. Verlag von Max Altmann, Leipzig (1910)
5. The Gospel according to Philip: Internet Archive. https://web.archive.org/web/20050212013152/http://www.metalog.org/files/philip1.html. Accessed 30 June 2020
6. St. Athanasius: On the Incarnation. De Incarnatione Verbi Dei. 54, 3: PG 25, 192B. In: Coptic Orthodox Church (1999). http://www.copticchurch.net/topics/theology/incarnation_st_athanasius.pdf. Accessed 30 June 2020
7. Weor, S.A. (eds., trans.): The Gnostic Bible: The Pistis Sophia Unveiled. Glorian Publishing, ISBN: 9781934206645 (2011). https://books.google.com.ua/books?id=Tex5bsdIZ6MC&printsec=frontcover&dq=The+Gnostic+Bible:+The+Pistis+Sophia. Accessed 30 June 2020
8. Rajneesh BS (1981) Belyy lotos (The White Lotus). Dukhovnoye znaniye, Moscow
9. Virilio, P.: The Vision Machine. In: Rose, J. (trans.). Indiana University, Indiana (1994)
10. Dunkan, A.: The Dance of the Future. Theatre Arts Books, New York (1928). http://www.mccc.edu/pdf/vpa228/the%20dancer%20of%20the%20future%20. Accessed 30 June 2020

Natalya Shelkovaya is the author of "Introduction to Religious Studies" (Russia, 2007), co-author of "Russia Abroad: An Anthology of Contemporary Philosophical Thought" (Boston, 2018) and textbooks for universities, "Religions of the World" (Ukraine, 2000), in addition to more than 150 scholarly papers published in Belarus, Moldova, Japan, Russia, the United States, and Ukraine. Her articles include: "Luther—Buddha of the West" (2003), "Mystical Life" (2004), "Cultures of the West and East: A Paradigm Shift" (2012), "From homo sacralis to Hi-Tech homo" (2013), "Frierdich Nietzsche on the Way Back to Oneself" (2015), "Erotology of Nikolai Berdiaev" (2015), "Spirituality in Circumstances of Auschwitz" (2016), "Meetings with V. Rozanov. Thoughts on sex. Sex and Androgynism" (2017), "East and West: The Culture of Heart and the Culture of Mind" (2017), "Religion and Spirituality" (in collaboration with Mikhail Sergeev, 2017), "Meetings with V. Rozanov. The Essence of Christianity" (2017), "Massivization and barbarism as phenomena of modern culture" (2019).

31 Integrated Science 2050: Transdisciplinarity

Simin Seyedpour, Ginn Assibey Bonsu, Carlos R. Rojas-garcia, Helen Kopnina, Marta Jaskulak, Mari Nieves Vergara, Derya Yılmaz, Alessandro Melis, Ehsan Keramati Niaragh, Mehdi Azizkhani, Clara Vasconcelos, Mohammad Sufian Badar, Armin Grunwald, Orlando Gomes, Dušan Marković, Guillermina Jasso, Aarne Mämmelä, Miloš Milošević, Silvia Sivera-Bello, Valentina Cauda, Emad AbouElgheit, Julie B. Ealy, Laura Stevens, Dario Dalla Vedova, Hideaki Yanagisawa, C. P. Hertogh, Natalya Shelkovaya, Ruth Oriama, Surapati Pramanik, and Nima Rezaei

Transdisciplinary researchers appear more and more as hope-makers.

Basarab 1996.

S. Seyedpour · N. Rezaei
Research Center for Immunodeficiencies, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

G. A. Bonsu
Department of Applied Design, Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Cape Town, Ghana

C. R. Rojas-garcia
Seafood Sustainable Farming, Research and Consultancy, Mexico City, USA

H. Kopnina
The Hague University of Applied Sciences, Hague, Netherlands
e-mail: h.kopnina@hhs.nl

M. Jaskulak
Laboratoire de Génie Civil Et Géo-Environnement, Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-FranceULR4515 - LGCgE, F-59000 Lille, France
e-mail: marta.jakskulak@univ-lille.fr

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_31

M. Jaskulak
Institute of Environmental Engineering, Czestochowa University of Technology, Czestochowa, Poland

M. N. Vergara
Department of Contemporary Art History, Fine Arts School, Complutense University of Madrid, Madrid, Spain

D. Yılmaz
Faculty of Languages, History and Geography, Ankara University, Ankara, Turkey
e-mail: d.yilmaz@ankara.edu.tr

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Ankara, Turkey

A. Melis
University of Portsmouth (UK), Portsmouth, UK
e-mail: Alessandro.melis@port.ac.uk

E. K. Niaragh
Department of Fine Arts, University of Tehran, Tehran, Iran

E. K. Niaragh · N. Rezaei (✉)
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Tehran, Iran
e-mail: rezaei_nima@tums.ac.ir; Rezaei_nima@tums.ac.ir

M. Azizkhani
Texas A&M University System, College Station, TX, USA
e-mail: m.azizkhani@tamu.edu; mehdiazizkhan@gmail.com

C. Vasconcelos
Interdisciplinary Centre of Marine and Environmental Research of the University of Porto; Departamento de GeociênciasAmbiente e Ordenamento Do Território/Unidade de Ensino das CiênciasFaculdade de Ciências, Universidade Do Porto, Rua Do Campo Alegre S/N, 4169-007 Porto, Portugal
e-mail: csvascon@fc.up.pt

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Porto, Portugal

M. S. Badar
Senior Teaching Faculty, DepartmentofBioengineering, University of California, Riverside, CA, USA
e-mail: director_academic@salfiainstitute.edu.in; msbadar@engr.ucr.edu

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Riverside, CA, USA

Director (Academic), Salfia Paramedical Institute, Darbhanga, India

A. Grunwald
Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Karlstr. 11, 76133 Karlsruhe, Germany
e-mail: armin.grunwald@kit.edu

O. Gomes
Lisbon Accounting and Business School, Lisbon Polytechnic Institute (ISCAL/IPL) & CEFAGE, ISCAL/IPL Research Center—July 2020, Evora, Portugal
e-mail: omgomes@iscal.ipl.pt

D. Marković
Faculty of Economics, Belgrade, Serbia
e-mail: dusan.markovic@ekof.bg.ac.rs

D. Marković · M. Milošević
Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Belgrade, Serbia

G. Jasso
Department of Sociology, New York University, New York, NY 10012-9605, USA
e-mail: gj1@nyu.edu

A. Mämmelä
VTT Technical Research Centre of Finland, Oulu, Finland
e-mail: aarne.mammela@vtt.fi

M. Milošević
Singidunum University, Belgrade, Serbia

S. Sivera-Bello
Universitat Oberta de Catalunya, Barcelona, Spain
e-mail: ssivera@uoc.edu

V. Cauda
Department of Applied Science and Technology, Politecnico Di Torino Corso, Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: valentina.cauda@polito.it

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Turin, Italy

E. AbouElgheit
Ancell School of Business, Western Connecticut State University, Danbury, CT, USA
e-mail: abouelgheite@wcsu.edu

J. B. Ealy
Penn State Lehigh Valley, Center Valley, PA, USA

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Nesquehoning, PA, USA

L. Stevens
The Hague University of Applied Sciences, Hague, Netherlands

Delft University of Technology, Delft, Netherlands

D. D. Vedova
Institute of Sport Science Antonio Venerando, CONI—Italian National Olympic Committee, Largo Piero Gabrielli 1, 00197 Rome, Italy
e-mail: dario.dallavedova@gmail.com; dario.dallavedova@coni.it

H. Yanagisawa
Marelli Corporation, 132 Shin-nakano, Ora-cho, Ora-gun, Gunma 370-0612, Japan

USERN, hideaki34, Ora-gun, Gunme, Japan

C. P. Hertogh
School of Foreign Languages and Cultures, Chongqing University (重庆大学), Chongqing, China

Department of Philosophy and Moral Sciences, Free University Brussels (VUB), BrusselsEurope, Belgium

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Tehran, Iran

N. Shelkovaya
Volodymyr Dahl East Ukrainian National University, Severodonetsk 93406, Ukraine
e-mail: nvshelkovaya@ukr.net; shelkovaya@snu.edu.ua

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Severodonetsk, Ukraine

R. Oriama
University of Strasbourg, Strasbourg, France

Institute for Public Policy and Governance (SIPPG), Strathmore University, Nairobi, Kenya

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Strasbourg, France

S. Pramanik
Department of Mathematics, Nandalal Ghosh B.T. College, Bhatpara, India

N. Rezaei
Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

Children’s Medical Center Hospital, Dr. Qarib St, Keshavarz Blvd, 14194 Tehran, Iran

Summary

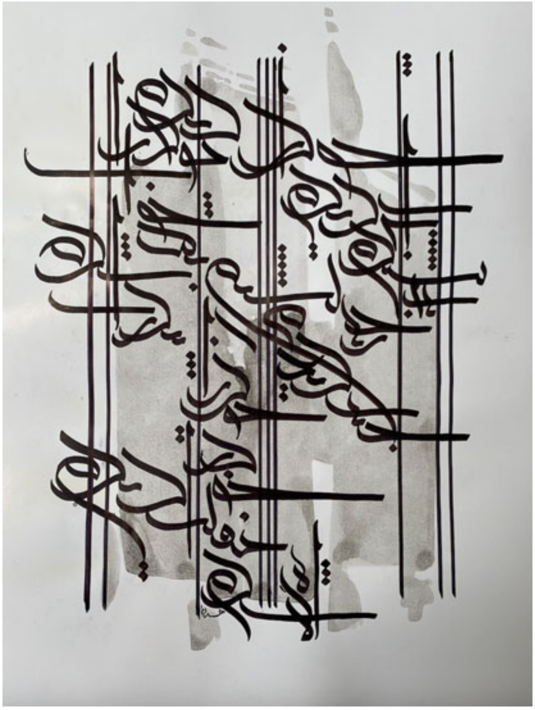
This chapter will briefly discuss the Integrated Science: transdisciplinary authors' perspective on how they see the future of their field in 30 years. The authors will briefly discuss their perspectives on the future of various topics, from sustainability to biomedical and economic challenges.

A sustainable Science for society

[Adapted with permission from the Medical Faculty Professor's Appreciation (Ghadre Ostad) Festival 2021, Artist: Simin Seyedpour].

code a

01110100 01110010 01100001 01101110 01110011 01100100 01101001 01100011 01101001 01110000 01101100 01101001 01101110 01100001 01110010 01111001



"I turned to water, saw myself as a mirage

I turned to sea, saw myself as a bubble

I become aware, found my own neglect

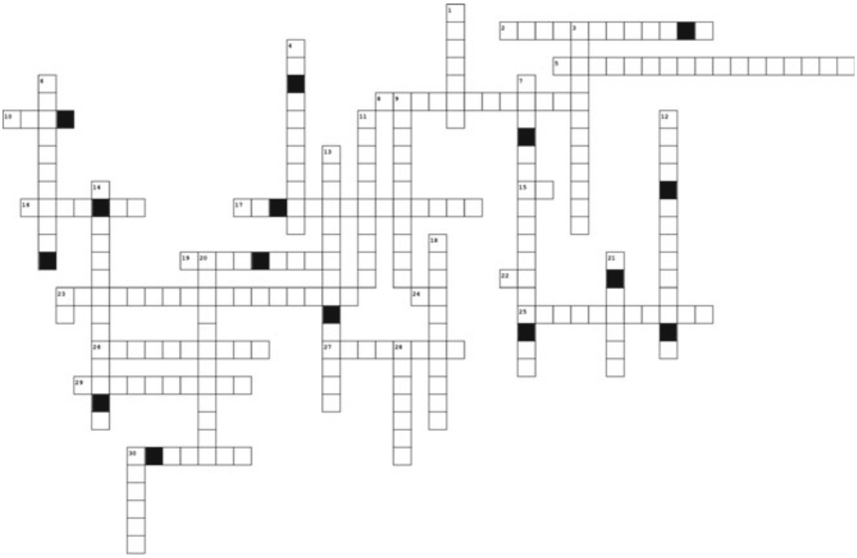
When I woke up, saw myself asleep"

[This is calligraphy of one of famous poets by Badakhshani, an Iranian poet; **Artist: Atoosa hedayati**]

code b

01010011 01110101 01110011 01110100 01100001 01101001 01101110 01100001 01100010 01101001 01101100 01101001 01110100 01111001

Name ___________________



Across

2. Word of chapter 4
5. Word of chapter 7
8. Word of chapter 9
10. Word of chapter 20
15. Word of chapter 24
16. Word of chapter 30
17. Word of chapter 11
19. Word of chapter 19
22. Word of chapter 5
23. Word of chapter 31a
24. Word of chapter 3
25. Word of chapter 1
26. Word of chapter 15
27. Word of chapter 18
29. Word of chapter 29
30. Word of chapter 6

Down

1. Word of chapter 28
3. Word of chapter 23
4. Word of chapter 12
6. Word of chapter 8
7. Word of chapter 16
9. Word of chapter 21
11. Word of chapter 10
12. Word of chapter 31 b
13. Word of chapter 13
14. Word of chapter 14
18. Word of chapter 26
20. Word of chapter 22
21. Word of chapter 2
23. Word of chapter 27
28. Word of chapter 17
30. Word of chapter 25

- ✓ There is a binary code for each chapter;
- ✓ There is a meaningful word for each chapter;
- ✓ The meaningful word of the chapter is made out of unsorted letters of the binary code of the chapter;
- ✓ There is a mystery phrase for the puzzle;
- ✓ The puzzle is filled in with the words of the chapters according to the clues provided below;
- ✓ The mystery phrase of the puzzle is made out of letters in black cells;
- ✓ Solve it, put your name on the puzzle, and send it to nimarezaei.usern@gmail.com

The puzzle of the Integrated Science: Transdisciplinary

1 Introduction

The rapid progress of science and technology such as artificial intelligence, data science, and robotics continuously alter our lifestyle, businesses, communications, health and impose the need for an upgraded educational system to qualify the next generations to overcome society's requirements and challenges [1–3].

Academic education remains the primary source of knowledge achievement in society, which is established based on disciplines. Traditional research mainly depends on institutions within academia that study world facts by classifying them into various disciplines. Experts and researchers investigate the answer for a specific question within the borders of that discipline [2, 3]. However, it will not solve questions that require a holistic and multidimensional approach, methodology, and solutions [3, 4].

The universities need to reshape themselves and provide opportunities where problem-solving will become the main focus and the common goal of diverse participants. Therefore, in order to meet future generations' educational demands, interdisciplinary research methodologies are needed to be more emphasized [3]. Transdisciplinary research provides a more profound integration throughout problem definition, investigation, and solution process [5].

In the previous chapters, the authors had discussed the different aspects of complex problems in each field and how transdisciplinary can offer a novel solution. In this chapter, we had asked the authors of the Integrated Science: transdisciplinary to prove their point of view on how they see the future of their field 30 years later.

2 Sustainability

As human civilization expanded, it started to consume and irreversibly harm its surrounding nature: air pollution, water inadequacy, animal and plant life distinctions, climate change, and global pandemics are some of the complex problems that affect humans and their ecosystems. Sustainability has been considered a problem-oriented approach that integrates various disciplines and stakeholders to maintain human civilization and nature's existence. Therefore, transdisciplinary is one of the best strategies to explore sustainability dimensions and recommend innovative solutions that consider the common good and participants' benefit [6].

We are at the dawn of an age where most societies have recognized—yet again—that their lives are solely dependent on their natural environment. Thus, any activity that destroys the environment puts their lives in danger and possibly to the point of extinction. The only means to see tomorrow is to look to the future by reducing 'defuturing' activities. Hybrid-local innovations fueled by global flows of ideas and creativity ensure the adoption of cosmopolitan localism is part of the future. Creative communities will be the heartbeat of our sustainable common futures, where locals promote the hybridity of design that sits between radical and transformative spaces. This design is birthed from global and local environments—but with less environmental impact. Therefore, the adaptation and usage of innovative technology will be censored, based on its environmental and non-human impact, not only on advantages and practical benefits to humans. By confronting defuturing activities, creative communities will be called to serve as gatekeepers for the present and the future through designing with the mandate of promoting cohabitation between humans and their environment (Ginn Assibey Bonsu 2021).

2.1 Resilience and Ecosystem-Based Management

In a rapidly changing world wanting any possible solution for a secure food supply, the citizens of future generations claiming their fundamental human right to food do not have to be a concern; they will require heredity golden farming tools to do so, their imagination is the limit. Sustainable shrimp farming is feasible anywhere on earth. Urban seafood farming in green cities must promote small horizontal economies, short value chains, and local supply chains without the need for a circular economy, certification, or traceability. We predict that in the next decade, freshly harvested shrimp will be available of fresh shrimp year-round in local food systems that will be antibiotics-free, hormones-free, pesticides-free, and pollution-free, that will taste good and be accepted by native inhabitants (Carlos r. Rojas-garcia 2021).

2.2 Ecodemocracy

While my background lies in social sciences and particularly in anthropology, I work in interdisciplinary fields within three main areas: sustainability, environmental education, and biological conservation. Regarding environmental issues such as climate change, biodiversity loss, pollution, and natural resource scarcity is challenging without engaging with practical as well as ethical aspects of various integrated disciplines. Examples of two chapters submitted to this volume together with multidisciplinary coauthors from fields as diverse as circular economy and anthropology (‘Cradle to Cradle and Cradle to Grave: Discussing the corporate case of eco-coffins’), and geography, pedagogical studies, conservation biology, and political science ('Ecodemocracy in Practice: Examples of Forestry and Food Production’) are examples of such integration. In the future, the fields related to sustainability, biological conservation, environmental politics, and related disciplines will hopefully further recognize both ethical and pragmatic implications of the intrinsic value of nature and individual non-human beings as an entity. Ways of considering root causes of unsustainability, including the somewhat uncomfortable issues associated with population growth and global expansion of material demands and lifestyles associated with increased production and consumption, will hopefully be developed. Limiting economies toward degrowth and re-orienting politics to include non-human representation would be a powerful solution. In order to learn to live, flourish, and share this planet with billions of other living beings and to guarantee brighter prospects for future generations will require expert input not just from scientists but also from the socially and morally engaged public, politicians, and other decision-makers (Helen Kopnina 2021).

3 Land Degradation

Land degradation is one of the challenges which can be discussed by sustainability. This phenomenon, which can be induced by human activities and natural disasters, leads to various socioeconomic problems such as food restriction and also treat agriculture, plant, and animal life. Land degradation poses a severe threat not only to our environment but also to human health. Decisions of managing natural resources determine our future well-being and the potential of our ecosystems to offer and design products and services. To balance the economy's demands, societies, climate, and the entire biosphere, we require a holistic and integrated approach. Hence, avoiding further land degradation is a crucial challenge that needs to be addressed worldwide. Moreover, it is apparent that land and water resource utilization are closely related. Consequently, they must be simultaneously managed with a consideration of biodiversity and the participation of the local communities in each region (Marta Jaskulak 2021).

3.1 Kaleidoscopic Image; Art and Science Intersection

It is always refreshing and challenging to think about the future. In the artistic research field, I guess that visual studies will become more and more needed, considering the virtual and ubiquitous world in which we are getting immersed at present. In my opinion, it is very possible to expect the kaleidoscopic image to be playing a key role within the next few years, both for scientific and artistic issues. At the same time, this configuration could be a significant intersection for trans-disciplinary collaboration between these fields and other disciplines, as is suggested in the title of the chapter. This thought is due to the references that I have been found recently, as well as during the development of my Ph.D. thesis on the kaleidoscopic image. It is a motif that has always been present in our imaginary perception, artistic expressions, and nature, from microscopic scale to the universe, although the term is getting highly associated with very innovative proposals in the last two decades. Who knows, perhaps in 2050, we can say that we live inside a kaleidoscopic multiconnected universe. There is no doubt that the discoveries in the next 30 years will be very astonishing (Mari Nieves Vergara 2021).

4 Landscape

The global dimension and magnitude of environmental crises lead us to believe that the paradigms of cities and, more generally, the human settlements, as we know them today, will be revolutionized in the next thirty years. In our opinion, the future of urban planning will be mainly oriented toward resilience as an instrument of adaptation to the continuous environmental changes due to the climate crisis. In this context, planning will be largely influenced by transdisciplinarity and, above all, by the indeterministic forms of city planning derived from the studies of the biology of evolution. The translation into the architecture of the exaptation will be a focus of these studies. This will imply that in the future, the redundancy and variability of the components of the city will be pivotal, as well as the diversity of the subjects participating in the design processes. Planners, therefore, will focus on design tools that can guarantee greater redundancy and flexibility in the design of cities, rather than aiming at a set of functions for future urban visions with specialized uses, such as those of the traditional 'zoning,' which, besides being proven wrong, did not allow a positive and sustainable development (Alessandro Melis 2021).

In the 2013 Paris conference, Bernard Lassus predicted that: "architecture, in order to save itself from formalism, and urbanization, to maintain itself against the dogma of programming, will have to turn to the landscape definition of space." To Western thought, the landscape is a by-product of modernity still defined on the foundation of seventeenth-century Cartesian Dualism. However, in twentieth-century thinkers such as Christian Norberg-Schulz, Christopher Alexander, Edward Relph, and Augustin Berque, who examined the phenomenon of "space," found out its dual indivisibility. Augustin Berque defines "landscape" as an objective–subjective phenomenon, in which the relationship between two components is neither $\vee$(or) and nor $\wedge$(and), but as an integrated phenomenon called 'objective–subjective'. Thus, we are welcoming the official emergence of a phenomenon which disclaims the four-century of the world of knowledge. Today, two definitions of space are simultaneously applied: Descartes' dualism in unscientific and non-academic circles, which recognizes space merely as "objective," and the European Union, which defines Landscape as "objective–subjective." The coming era will see the decline and collapse of the aforementioned Cartesian Dualism toward 'space' and accepting a compound 'totality' of the objective and subjective components as two aspects of a sole existence. Recognition of a new phenomenon as "whole," which indicates a combination of physical and metaphysical matter, will occur in the world ahead. "Landscape" will be the gateway through which modernity will be admitted into a new era when subjectivity will play an inescapable role in the valuation processes of human societies (Ehsan Keramati Niaragh 2021).

4.1 Biomimicry in Architecture

Biomimicry is considered to be one of the main tools in our transition from an industrial age to an ecological age between now and 2050 [7]. We are observing, and we will see, such a transition in the building industry, but this transition may not happen soon and effortlessly through biomimicry, mainly because the transdisciplinary collaboration needed for a biomimetic design can rarely happen in a building design and construction process. In a building design/construction team composed of stakeholders, contractors, clients, users, designers, engineers, and many other decision-makers, we rarely see physicists, scientists, biologists, and also experts who are competent in using a transdisciplinary language to communicate with other disciplines. While we are moving toward transdisciplinary education/learning in academia, the professional environment of the building design and construction industry will not be ready soon to digest this transdisciplinary transition. Justification of the required professional education, additional cost of design and construction, and the feasibility of any proposed method to replace an existing design/construction trend will play important roles in facilitating or hindering this transition. Currently, the architectural methodologies of transferring our discoveries in nature to the level of scientific emulation in reaching building designs and products are still vague and very conceptual. Therefore, early future efforts may focus on facilitating this translation of concepts to actual products grounded in a visible, systematic, repeatable, detailed, and scientific process. Otherwise, much of the scholarly work on biomimetic architecture will remain unutilized and forgotten. Technology will be a major component in this process, and we should expect to hear more about "biotechnology" rather than "biomimicry" in the future trends of architecture and construction (Mehdi Azizkhani 2021).

4.2 Geoethics

One of the existential challenges with which citizens in the twenty-first century must contend is coexisting peacefully with the environment. Most of the environmental hazards that put humankind at risk are related to the geoscience scientific discipline. Moreover, the top risk, global climate change, actually questions the existing ability of the life of humans on earth. The inspiring geoscience research in geoethics and its excellent findings provide significant tools for dealing with environmental problems. Intensive research should be required to identify which educational and communication strategies should be adopted to convey the geoscience value to society and to suggest effective ways of encouraging geoscience professors to move from traditional teaching toward a progressive teaching method and how to develop communication skills with laypeople who are not familiar with the scientific jargon and lacking in scientific background. Those studies should be supported by research that focuses on altering geoscientists' attitudes toward the implementation of geoethical values and their role in society. The recognition that geoscience literate societies are an existential need and the current gap between the geoscience community attitudes toward their social responsibilities and their limited skill for interacting with the public will put the geoethics as an important and respectable field of the geosciences, namely with the great effort of Geoscience Education to give a huge step forward this inclusion (Clara Vasconcelos 2021).

5 Economic, Technology and Industry

The social and economic challenges are interrelated with environmental protection. Economy, technology, and industry are some of the main dimensions of sustainable development goals, which are determined to provide a sustainable future.

As Yogi Berra once said, “It's tough to make predictions, especially about the future.” However, the way science is done will be able to overcome the existing social and technical barriers. Fundamentally, the science will become multidisciplinary. Users can access networks, supercomputers, data science, software, and other scientific and data facilities in an integrated environment. A big change that will occur is creating data facilities that provide well-curated and long-term data repository services and inter-disciplinary data analytics services. User facilities will be more automated than ever, so much so that the line between real-world synthesis and experimentation through automation and simulation will be blurred. A trained scientist will find data and workflows similar to how most people find information on the Internet. A user will be able to contribute data to data facilities using existing standards and tools that will check for data compliance and quality and automatically tag the data with the appropriate metadata. Artificial intelligence will become mainstream, helping users to write their workflows and finding data, enabling data fusion across datasets from different domains. The way that science is currently progressing, science and technology will merge in such a way that it will be easy to use, fix, understand its result/analysis. These advancements in technology will be feasible only when there is a complete integration of science (Mohammad Sufian Badar 2021).

Technology assessment has been a growing field now for almost a century. After its start in Western countries, TA spread over the globe and became an international and intercultural movement. Based on sound science and careful reflection, it succeeded to contribute to harvesting the fruits of the scientific and technological advancements in many areas to a full extent, e.g., for transforming energy supply systems to a sustainable and climate-friendly mode of operation as well as for embedding autonomous technologies into human life without endangering human dominion. Simultaneously, humankind learned from technology assessment how to avoid unintended side-effects of technology. This was possible because TA contributed to changing the mindsets of people toward a more responsible and sustainable behavior. A nice dream? Yes, it is a dream. But sometimes, dreams really can make a difference and considerably change reality. Let's go forward keeping this dream in mind (Armin Grunwald 2021).

5.1 Macroeconomics

Economics occupies a unique place in science. On the one hand, it is a social science dealing with people's behavior and their interaction, with all the unpredictability that such behavior and such interaction involve. Simultaneously, on the other hand, it has always aspired to be guided by unambiguous and uncontestable principles, very much like natural and life sciences. The balance between these goals has been achieved by designing simple equilibrium models based on the robust notion of rational expectations: Economic agents plan for the future makes complete and efficient use of the available information and without incurring systematic mistakes. As economic relations intensify and grow in complexity, new analysis tools must be associated with the arsenal of instruments that this science resorts to address, explain, and act upon economic phenomena. Accompanying a general tendency in science, economics will undoubtedly become more organic, procedural, and algorithmic, relying on its analysis no longer in simple analytical toy models but in extensive scale simulations among agents that are heterogeneous at a variety of levels (preferences, forecasting abilities, endowments, skills). In short, economics in general, and macroeconomics in particular, will embrace complexity as their central perspective upon reality (Orlando Gomes 2021).

5.2 Automotive Industry

Although it is tough to give a long-term forecast about the development of any vibrant industry, and the automotive industry is undoubtedly one of them, because it has been facing disruption for more than a decade, I believe that in the future, the focus will definitely be on electrification, digitalization, connectivity, and autonomous vehicles. Accordingly, software and content will make up the lion's share of the perceived value of vehicles. Additionally, the digitalization and development of autonomous vehicles will positively affect the development of the market of modern mobility services, and some consumers will give up owning a car. Auto companies will invest in new technologies, like flying cars, which will probably supplant contemporary products. Emerging markets will be the main drivers of global demand. Regional differences in customer preferences will be lower than today, but they will still exist, and auto companies will have to adapt the marketing mix to them. Digitalization and artificial intelligence will radically change the organization of production. The industry will be less labor-intensive than today, but the technological intensity of the industry will increase. I suppose that competition will be fierce, and today pioneers and fast followers will lead the industry. Few emerging markets multinationals will be strong competitors in the industry. The source of competitive advantage in the industry will be lean thinking, business flexibility, and the ability to efficiently organize a strategic business network with innovative suppliers and incorporate innovative startups and even competitors in the network (Dušan Marković 2021).

6 Science, Mathematics, and Art Unity

It is easier to imagine the future than to give it a timetable. Too many things can interfere. But sooner or later, the road to the essential unity of knowledge will have three signposts:

1. Rededication to universal human development and universal human rights. The growth of knowledge requires full development of everyone's potential, and human development requires human rights. There is little doubt that the world is on course, however slow or opaque it may sometimes appear.
2. Claiming mathematics, a human birthright.
3. The world will awaken from the dark night of math anxiety and will learn how to impede its transmission.
4. Mathematics will not only accelerate the growth of knowledge and promote works of art and imagination-via its unparalleled ability to develop abstract thinking-but also it will bring joy and beauty.
5. Building a chain of conference centers, on islands belonging to the planet. These will have two features. First, they will welcome everyone and besides providing meeting space will host concerts and public lectures and have museums. So that human development and the growth of knowledge can flourish, neither visa nor bank account will be required to enter the conference centers. As Leonard Bernstein dreamed, birth in the human tribe will confer the special passport of the citizen of the world. Second, because staffing a conference center requires the broadest set of skills, the conference centers will be a place where anyone, of any age, can acquire a new skill or gain experience (Guillermina Jasso 2021).

7 Systems View in Engineering Research

In 2020–2050, technologies approach the fundamental limits of nature, and complexity becomes a major problem. Solutions to complex problems are found using systems view. The developed large-scale smart systems can solve the tragedy of commons with subsidiarity: Only essential coordination of the commons is performed centrally, and the rest is decided locally. Inviting more disciplines to the research efforts enables developing solutions for society. The disciplines will jointly advance the systems view in intractable nonlinear systems. They will also develop smart systems for societal systems, solve open problems in engineering, and refine experimental research methods. Smart systems are extended to systems supporting the decision-making of society and individuals. These data-driven systems solve problems related to overpopulation, overuse of resources and pollution with regulation that is based on international contracts. The systems observe the environment and suggest decisions to decision-makers at different levels from individuals and municipalities to states and international organizations preparing agreements between states. In addition, society and individuals have given consent for smart systems to provide autonomously a rich set of local services. These systems fulfill the needs of society and the individuals while using resources efficiently. Feedback produces insight on the effect on the decisions selected by the society, the individuals, and the autonomous systems—hence they all learn over time to make better decisions. Internationally built incentive systems attract also companies to develop smart systems that improve sustainability. Transdisciplinary efforts using systems view develop the smart systems further (Aarne Mämmelä 2021).

7.1 The World Big Challenges Faced by Nanoscience

Constructive human cooperation is essential to try to solve our world's biggest problems and challenges. Science and engineering can certainly contribute, adopting a broad and multidisciplinary approach as well as cutting-edge techniques to promote innovation in various fields of science and technology. Nanotechnology deals with the creation and manipulation of matter on a scale ranging from 1 to 100 nm. This scientific field, emerging already in the late'50 s, has so far revolutionized the comprehension of matter and opened the door to a plethora of novel, unprecedented and intriguing applications and solutions to current problems. The role of nanotechnology in facing the current world's challenges is acquiring increasing importance. Nanotechnological solutions to the most urgent World Challenges of our world are totally or in part contributing with other technologies to these challenges' solutions. Some of the possible solutions are already in place; others are under investigation at the research levels. Even others still require a lot of work and commitment by academic, industrial, and even policymakers to become reliable solutions. The hope is that in the near future, nanotechnology researchers with a broad, visionary, and multidisciplinary approach will become industrially viable solutions to the world's challenges and improve our quality of life, contributing to a better world (Valentina Cauda 2021).

8 Media and Viral Marketing

As a product of technology, media plays a crucial role in providing interaction among various disciplinary and exposing various partners to the dimensions of complex problems.

Augmented reality (AR) will alter marketing forever. Before 2030, customers may not be able to differentiate what is real from what is not within an AR experience. Future marketing content will serve multiple customer realities (i.e., real, enhanced, favorable, and mixed). Marketers will have to understand and influence consumer experience across multiple realities (Emad AbouElgheit, 2021).

In the next 30 years, we should expect an even greater impact of new media on both the media and social reality through further massification of their production and consumption and intensification of all today's visible directions of transformation of both the media and society. The day is not far off when technological progress will enable the overcoming of screen technology and exceeding the "immutable distance," i.e., the literal entry of the media into the subject as well as the subject into the media discourse. A virtual reality medium conceived on post-screen image technologies will acquire the properties of simulated reality while the boundary between media and real-life will no longer exist. All media and art forms known today will be integrated into one big metamedia. Everyday human experience will be mediated and shaped by different media content and technologies. Today's level of technological development in relation to the next 30 years will be marked by the term "stone digital age".

This development of technology and the scope of its use will bring radical changes both on a personal and social level. The boundaries between media science, human science, and social science will be erased. The research of such a world and man will be matched exclusively by a new metascience that will integrate all the research methods and approaches known to date (Miloš Milošević 2021).

Social media has an essential role in marketing as it can provide a shortcut for companies to advertise and introduce their products to customers without investing in traditional marketing. Boase and Wellman carried out a comparative study of biological, computer, and marketing viruses to understand how network structures affect their spread: how many people might be affected and at what speed. For all three viruses, frequent contact increased the probability of contagion, but only marketing viruses were welcomed, as they provide information or an opportunity to feel socially accepted and trendy [8].

This type of multidisciplinary study will make even more sense in the near future because, with the rise of fake news and hate speech predominating in certain social networks, we need to understand the mutation mechanisms that enable the viruses to survive. In the same way, as fact-checking tools are now used, people might turn to ad-checking to neutralize ambiguity in advertising authorship and to shine a light on the credibility of influencers. With regard to brands, the inclusion of watermarking, as part of the advertising virus genome, could provide a solution to the challenge of the measurement and effectiveness of communication (Silvia Sivera-Bello 2021).

9 Archeology

The emergence of artificial intelligence, data science, and advanced imaging technologies greatly impressed biomedical research, which is another critical issue discussed by transdisciplinary methods. Moreover, the emergence of these technologies had dramatically influenced biomedical research and health systems and found significant applications in related sciences such as archeology and crime science.

Archeology in the future will undoubtedly be exciting with newly developed excavation methods. In parallel with the development of technology, robots in some works may be on the agenda. For instance, a robot can easily replace today's workers for an excavation with a sensitive sensor. Moreover, photoshoots can be done with photography robots offering multiple options. Assuming that computer and robotic technology will advance, a robot, a great excavation assistant knowing and remembering everything, can work with us. In addition, drone technology will provide new scanning methods, meaning new camera systems.

Contributions of today's science branches to archeology which is a past science are increasing day by day. The relation between archeology and archeological sciences is long and will lead to new unique studies in the future. In parallel with developments in space studies, discoveries may emerge in archeoastronomy. I think the most significant discoveries will be about the melting of glaciers in the future. New knowledge of human history will be obtained solidly from the frozen lands. In parallel with technology development, the developing archeological sciences will continue to be more integrated into archeology.

On the other hand, discoveries under the ground that are not yet known will be exciting. DNA archeology will continue to reveal the identities of owners of particular cultures, of which we cannot be sure nowadays. It will be possible to take a virtual tour through the streets of a digitally crafted three-dimensional ancient city or even chat with some ancient personalities. Imagine going to the ancient city of Assos and having a philosophical conversation with Aristotle. With the digital age, the future of archeology will be exciting (Derya Yılmaz 2021).

10 Biomedical Sciences

In 2050, the Age of Biology has emerged, and the field of biomimicry is integrated into higher education around the world as a major program to educate professionals on how to incorporate the knowledge found in nature into design solutions of all scales. The climate crisis has been overcome, human poverty is non-existent, and energy, food, and material resources are readily available, mainly because of our ability to translate biological strategies into systematic, regenerative design solutions that supply ecosystem services. Businesses are eager for biomimicry professionals to enter their workforce to educate others on how to improve their ecosystems thinking in a manner that is inspiring, beautiful, and iteratively circular, helping others to find their own potential within the system they exist in. Social innovation is thriving because of the shared knowledge and skills; humans are finally considered a valuable Earth organism, fully integrated within their place in nature. ‘Cooperation’ is the fashionable, fundamental favored movement, becoming the norm. And finally, there is no need for labels such as fair trade, organic, biomimetic, or sustainable because, in 2050, we understand how the planet works, and at that point, it will be what and how we do things naturally (Laura Stevens 2021).

One of the paramount hopes of the author of the chapter, “The Artistic and Scientific Nature of Proteins: A Historical Perspective,” is that there will be significant advances in our knowledge and understanding that develops from computational software applications allowing many protein structures to be determined; drugs can be designed that will have fewer or no side effects. Perhaps advances in computational software will occur so that an initial molecule chosen for drug synthesis will be the drug that works, and the time to market will be shortened as well as a dramatic decrease in expenses associated with drug development and cost. In addition, substantial new knowledge about the individual genetic makeup of each person should enhance the field of pharmacogenetics so that drug dosage can be tailored to an individual’s metabolism and genetic profile. This will help prevent drug under- and over-doses, especially because one size does not fit all.

Indeed, the hope is that there will always be those enthralled with the beauty of protein structure, such as artists whose scientific knowledge and understanding result in a sculpture or painting that transcends time. But we also hope that there will always be scientists who want to be ‘workers’ and who strive for scientific advances in protein structure (Ealy 2021).

Basic technologies, inputs, and demands form an integral part of the building blocks in a sectoral system of innovation. Biosimilars development has emerged as an opportunity for incremental innovation in the heath bioeconomy, and is set to affect the political, institutional, and niche-level arrangement of economies inclined to cleverly valorize their biological resources. The bioeconomy can be leveraged to provide a competitive advantage for economies and firms involved in the biosimilar value chain.

Economies like Kenya in the next three decades are set to benefit from the expiration of biologic patents, to secure opportunities for universal health coverage and national competitiveness. Emerging economies as seen in the case of Asia and Africa are already engaging realignment transformation pathway of the multi-level perspective where multiple niche innovations are set to coexist and compete, until only one innovation remains, which then replaces the old technologies. Healthcare and regulatory pathways will evolve from being barriers to accommodating the production of advanced therapies such as biosimilars, regenerative and stratified medicine.

The diffusion of knowledge related to the new biology has accompanied borderless global opportunities like the new biology and biosimilar development. This diffusion has occurred to the extent that reverse innovation is disrupting previous positions for internationalization and technological advancement. The same is observed for borderless challenges such as access to health care which will continue to attract global efforts for solutions (Ruth Oriama 2021).

On a larger scale, health-related activities such as sport can be a subject of transdisciplinary research regarding their complexity and unpredictable effects on each individual.

Playing a game or a sport is a voluntary attempt to overcome unnecessary obstacles using rules to forbid the most efficient means [9]. Performance is a multidimensional construct made of a complex mix of athletes, rules, equipment, and environment intersecting physics, chemistry, and biology. Gaining the podium is a matter of detail. So, room for improvement will always exist. Athletes are the less and the more variable element of the system. The less varying because the human body will reasonably remain the same, the more developments in different fields are unpredictable. Sports rules, equipment, and records will depend on how the show and the business will develop. Each body is unique, making sensations, balance, and fatigue strictly personal. Sports know-how is widely distributed, partially tacit, and non-codified. We deal with complex, open, and adaptive systems. Due to countless interactions between parts, emerging phenomena, redundancy of degrees of freedom, we cannot control or predict them beyond certain limits. We can only disturb the system by seeing what happens in a challenging process of trial and error. The border between experimental, repeatable, falsifiable hard sciences and theoretical, historical soft ones is shaded in sports. Science of complexity brought several answers and opened new frontiers which can be explored with increasingly heterogeneous expertise. The integration of different pieces of knowledge is necessary, especially at the top level. Women and men do sport with their emotions, passions, weaknesses. It will change and evolve to remain that exciting and unpredictable phenomenon we love (Dario Dalla Vedova 2021).

11 Transdisciplinary in Philosophy and Science

Leibniz said, ˵I hold time to be an order of successions.'

In current science, an absolute atomic clock presupposes the presence of absolute length, and an atomic absolute telemeter presupposes the presence of absolute time. Therefore, a ˵chicken and egg' contradiction exists in them. Furthermore, the arrow of time is explained with the second law of thermodynamics. However, the second law of thermodynamics can't adapt to energy without matter. Time can be defined with the continuous covariant relation of energy according to chaos theory. The arrow of time can be explained with chaos theory, too. All academic fields must be corrected with the new definition of time because all-natural phenomena, including human thinking, obey chaos theory. Regarding natural phenomena, all theories without considering time-based change are insufficient. Current academic disciplines and religion are divided into many fields. Given that a subject's relation to another is pretermitted with analysis, abuse and war may arise. Therefore, research and academic disciplines without decreasing total entropy may harm living creatures such as pollution and war. In contrast, human thinking unified with chaos theory greatly benefits living creatures because total entropy decreases in all academic fields. Thirty years later, decreasing total entropy over time will become a standard in all academic fields, including each religion. Because human thinking will be unified in the correct direction, abuse and war caused by separation will decrease in each field (Hideaki Yanagisawa 2021).

The Eastern worldview of cyclic world development appeals to me. Therefore, I do not share the point of view of post-structuralists about the "death of philosophy" and "death of a person" (M. Foucault) after the "death of God" F. Nietzsche is thought to have proclaimed (although Nietzsche emphasized not the death of God, but the killing of God by man, His death for man).

Although man died, however, not in the twentieth century, but immediately after the Fall from grace, that is, as a result of separation from God, as God himself warned them about, still… It is *obligatory* to die in order to rise again. Consequently, following the development in a cycle, a person must inevitably resurrect as a Person, that is, become again the God-like, but not the beast- or machine-like creature that they have become today.

All mystics of all religions remind of the "return to the self" of a person as a return to God. Philosophy will also "return to itself" pristine as a love of wisdom, as a search for Truth with a capital letter. The split of a person and the dichotomy of the worldview will be replaced by a holistic worldview in which philosophy, religion, science, and all philosophical, religious, and scientific directions will not only stop the meaningless confrontation but form an organic complementary and mutually developing unity, and thereby gain the ability to transcend from the physical world and its "man-made realities" into metaphysical, for only "from a height," as F. Nietzsche put it, it is possible to comprehend what is happening below, on Earth (Natalya Shelkovaya 2021).

For the next thirty years, we expect that two conditions to twenty-first-century philosophy and science will keep on being materialized in science and society, environmental Pragmaticism, and global cross-culturalism [10–12].

We would like to add two postmodern conditions to the twentieth-century philosophy of Science—global cross-culturalism and environmental pragmaticism. Concerning the former, see, e.g., Douglas Hofstadter's cross-cultural hypothesis in this paper [e.g., koans, Penrose figures, AI as examples of 'feedback loops'—prerequisite that science draws on a corpus of at least two globally different cross-cultural sets of data]. Concerning the latter, it wants to eliminate possible risks of damage to human health and natural environments from (societal applications of) science.

We can summarize these maxims in addition to Karl Popper's formula on the growth of knowledge, the progress of science ('…fundamental evolutionary sequence of events… P1 → TS → EE → P2'—Popper 1979, 243), modifying it into the progress of science and society:

P1 → TS → EE/EP GC → P2.

where P stands for Problem, TS for Tentative Solutions, EE for Error Elimination, EP for Environmental Pragmaticism, and GC for Global Cross-culturalism. The forward slash at EE wants to allow for twenty-first-century global cross-cultural and environmental criticisms—as to save Planet Earth.

> When the moment comes to choose between economic growth and ecological stability, politicians, CEOs, and voters almost always prefer growth. In the twenty-first century, we shall have to do better if we are to avoid catastrophe.

However, in our view, green policies and economic growth do not necessarily exclude, but possibly enhance each other, and we need to change to green technologies etc. as soon as possible, e.g., by the development of renewable energy instead of fossil fuels (Hertogh 2021).

Transdisciplinary will provide a systemic approach for knowledge integration to solve the world's challenges by taking advantage of cutting-edge technology in various fields. In brief, transdisciplinary research aims to comprehend the dimensions and complexity of each problem, respect various perspectives on a problem from academic and non-academic organizations, interrelate the theoretical and practical knowledge, and acknowledge the conflicts and recommend solutions regarding the common good [3, 13].

12 Conclusion

Integrated science refers to a science that integrates several different disciplines such as biology, microbiology, ecology, physics, chemistry, Earth science, and astronomy [14]. Integrated science has been implemented in different countries on all continents, but it did not appear as successful as expected [15].

Researches in integrated science have been conducted in various countries. Winarno et al. (2021) presented a comprehensive and critical review on the implementation of an integrated science curriculum. According to the United Nations report, the world population is expected to attain 9.7 billion in 2050. The world economy could be more than double in size by 2050, but poverty will sharply increase in sub-Saharan Africa, Middle East, and Latin America around 2050 that will create several problems in the education sector. So integrated science education will have to deal with inequality, uncertainty, inconsistency, and indeterminacy around 2050. The COVID-19 outbreak made a significant impact on educational systems globally and led to the near-total closures of educational institutions of all age groups. The curriculum of integrated science around 2050 for higher education will have to deal with the consequences of COVID-19 and help to prevent future recurrences. To do that, integration of microbiology, molecular pathology including immunology, and epidemiology is required. Around 2050, an integrated science curriculum will reflect a science, technology, engineering, and mathematics (STEAM) curriculum to meet the demand of society. Around 2050, it is expected that machines with autonomous intelligence and self-evolution capabilities will be developed, and integrated science for higher education appear will contribute to developing such machines. Important research and development in integrated science around 2050 are expected to include (Surapati Pramanik 2021):

- Integrated science learning media based on science-edutainment
- Integrated science learning tools using a humanistic approach assisted with props
- Studies reviewing journal articles on integrated science
- Evaluation of integrated science education.
- Instructional objectives and integrating techniques of integrated science courses.

Core messages

- Transdisciplinary research will significantly alter human perspective toward the global challenges and potentially resolve socioeconomic conflicts while appreciating sustainability.
- Media will provide a unique platform for know-how exchange and knowledge integration.

References

1. Rampisela DA (2020) Transdisciplinary Research Bridges the Gap between Science and Pastoral Services. Journal Jaffray 18:95–105
2. Hadorn GH, Hoffmann-Riem H, Biber-Klemm S, Grossenbacher-Mansuy W, Joye D, Pohl C, Wiesmann U, Zemp E (2008) Handbook of transdisciplinary research, vol 10. Springer
3. Guimarães MH, Pohl C, Bina O, Varanda M (2019) Who is doing inter-and transdisciplinary research, and why? An empirical study of motivations, attitudes, skills, and behaviours. Futures 112:102441
4. Wernli D, Darbellay F (2016) Interdisciplinarity and the 21st century University
5. Bernstein JH (2015) Transdisciplinarity: A review of its origins, development, and current issues
6. Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ (2012) Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain Sci 7:25–43
7. Abou-Houly H (2010) Investigation of flow through and around the macrotermes michaelseni termite mound skin. Loughborough University
8. Boase J, Wellman B (2001) A plague of viruses: biological, computer and marketing. Curr Sociol 49:39–55
9. Paddick RJ (1979) The grasshopper: games, life and Utopia. By Bernard suits. Toronto, University of Toronto Press 1978. J Philos Sport 6:73–78
10. Hertogh C (2015) Semantics of thought experiments'. PhD diss., VUB Brussels
11. Hertogh C (2020) Goodman's Grue? Relativized Pluralism and Paradigmatic Thought Experiment (Whack). Humanit Bull 3:64–103
12. Hertogh C (2018) Logical analyses of vipassanā meditation. Contemp Buddhism 19:251–274

13. Pohl C (2011) What is progress in transdisciplinary research? Futures 43:618–626
14. Hewitt PG, Lyons S, Suchocki J, Yeh J (2014) Conceptual integrated science. Pearson Education Limited
15. Winarno N, Rusdiana D, Susilowati E, Afifah RMA (2020) Implementation of integrated science curriculum: a critical review of the literature. J Educ Gifted Young Sci 8:795–817

Correction to: Chaos Theoretical Explanation to Each Development of Evolution Theory, Psychology, Physics, and Philosophy

Hideaki Yanagisawa

Correction to:
Chapter 2 in: N. Rezaei (ed.), Transdisciplinarity, Integrated Science 5, https://doi.org/10.1007/978-3-030-94651-7_2

In Chapter 2, the removal of the graphical abstract and equations has been corrected. The correction chapter and book have been updated with the changes.

The updated original version of this chapter can be found at
https://doi.org/10.1007/978-3-030-94651-7_2

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_32

Index

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7