

Open Peer Review on Qeios

Measuring Complexity using Information

Klaus Jaffe¹

1 Universidad Simón Bolivar

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Measuring complexity in multidimensional systems with high degrees of freedom and a variety of types of information, remains an important challenge. Complexity of a system is related to the number and variety of components, the number and type of interactions among them, the degree of redundancy, and the degrees of freedom of the system. Examples show that different disciplines of science converge in complexity measures for low and high dimensional problems. For low dimensional systems, such as coded strings of symbols (text, computer code, DNA, RNA, proteins, music), Shannon's Information Entropy (expected amount of *information* in an event drawn from a given distribution) and Kolmogorov's Algorithmic Complexity (the length of the shortest algorithm that produces the object as output), are used for quantitative measurements of complexity. For systems with more dimensions (ecosystems, brains, social groupings), network science provides better tools for that purpose. For complex highly multidimensional systems, none of the former methods are useful. Useful Information Φ, as proposed by Infodynamics, can be related to complexity. It can be quantified by measuring the thermodynamic Free Energy F and/or useful Work it produces. Complexity measured as Total Information I, can then be defined as the information of the system, that includesΦ, useless information or Noise N, and Redundant Information R. Measuring one or more of these variables allows quantifying and classifying complexity.

Klaus Jaffe

Universidad Simón Bolívar, Caracas

Keywords: Thermodynamics, Infodynamics, Complexity, Information.

Introduction

The aim of this paper is to help create bridges between disciplines that allow researchers to share tools to advance in the building of a unified robust science of complexity. One such tool is the measurement of complexity. Attempts to measure



complexity are not new [1][2][3][4]. They are based upon the assumption that complexity is proportional to the number elements and diversity of a system, or of symbols needed to code a given information and eventually to the relationships among them. As recognized by Wikipedia, the term Complexity is generally used to characterize something with many parts that interact with each other in multiple ways, culminating in a higher order of emergence greater than the sum of its parts. The study of these complex linkages at various scales is the main goal of complex systems theory. The intuitive criterion of complexity is related the number of parts of a system and the amount of connections between them existed [5].

Several approaches to characterizing complexity have been used in different sciences^[6] Among them, relating complexity to information has a long tradition ^{[7][8][9][10]}. There is however no unambiguous widely accepted strict definition of complexity. In <u>computer science</u> and <u>mathematics</u>, the Kolmogorov complexity ^[11] of an object, such as a piece of text, is the length of a shortest <u>computer program</u> that produces the object as output. It is a measure of the computational resources needed to specify the object. In multidimensional systems, this definition is of little use as the Kolmogorov complexity becomes incomputable. Fisher's ^[12] Information and Shannon's Information Entropy ^[13] estimate complexity by the information content of a given system, however his method becomes algorithmically incomputable in systems with high dimensions. These methods have large limitations for studying phenomena which emerge from a collection of interacting objects ^[14], especially if high dimensional systems are being studied.

"The science of complexity is based on a new way of thinking that stands in sharp contrast to the philosophy underlying Newtonian science, which is based on reductionism, determinism, and objective knowledge... Although different philosophers, and in particular the postmodernists, have voiced similar ideas, the paradigm of complexity still needs to be fully assimilated by philosophy" [15]. Fortunately, in natural sciences, these concepts seem to be more assimilated, though in different disciplines they have different flavors.

Complexity in different disciplines



Discipline	Object	Measured through	Example
Quantum Mechanics	Subatomic particles	Taxonomic complexity	Standard Model
Inorganic Chemistry	Atoms	Taxonomic complexity	Periodic Table
Organic Chemistry	Molecules	Taxonomic complexity	MS-AI
Biology	Organisms	Taxonomic complexity	Tree of Life
Genetics	DNA, RNA, Proteins	Coded strings	Genome
Language, Music	Scripts	Coded strings	Texts, Songs
Computing	Bits, Qbits	Coded strings	Programs
Ecology	Systems, organisms	Networks	Ecological webs
Social sciences	Humans	Networks	Constitutions, Society
Economics	Values Industries	Coded strings Networks	Stock market Economic Complexity
Physics	Mater, energy	Degrees of freedom	Emergence
Infodynamics	Energy, Information	Work achieved	Free Energy
Artificial Intelligence	Models	Work achieved	Al programs

Taxonomic Complexity: The simplest way of measuring complexity is by counting the parts or components of a system. The examples given in the table include the Standard Model of particle physics. This model is a descriptor of the complexity of the quantum world enumerating all known subatomic particles. These 16+ particles explaining 3 fundamental forces of physics allow us to express quantitatively the complexity of the results of interactions between one or more of these subatomic particles. Analogously, the Periodic Table of Chemistry lists 118 known chemical elements. They are ordered in increasing complexity related to their number of protons and electrons. The combination of any number of different elements produces an immense universe of possible molecules whose complexity can be assessed by the number of different types of atoms forming the molecule or by the complexity of different manifestations of the molecule [16]. The combinatorial complexity is so great that artificial intelligence (AI) has been called to compare their characteristics [17][18]. Similarly, the complexity of life on earth can be graded qualitatively in degrees of complexity in accordance to the complexity of the organism and their position in the universal phylogeny. The variables affecting complexity in living systems, however, have not been tamed so as to produce meaningful measures on complexity. I refer to these types of complexity as Taxonomic Complexity.

Algorithmic Complexity: Another type of complexity can be referred to as information coded in strings of symbols. This is the case of Genetics, Animal and Human Language, Music and Computing. In all these cases, information can be represented as strings of coded characters such as bits, qbits, letters, numbers, nucleotides or amino acids. As the representation uses only one dimension in the sequence, and the coded characters are finite, analytical scalar measures of complexity, even very sophisticated ones, are possible ^[19]. Following the insight of Shannon, information entropy



content can serve as a efficient measure of complexity ^[20] and can be applied to music ^[21] literature ^[22] genomes ^[23] and computer language ^[24]. Alternatively, Kolmogorov Complexity continues to be a widely used method ^[25] for sting codes. Other methods are based on order that eventually can be reduced to information entropy ^[26], or are based on specific physical properties ^[27]. Complexity analysis of series of marker data ^[28] show that time series of values of commodities in financial markets contain complex structures that help understand fundamental characteristics of the markets ^[29].

Networks: Not all natural phenomena can be reduced to strings of code. For two or more dimensions, network science has developed a series of complexity measures ^[30]. These measures of "aggregate complexity" are sometimes extension of computer complexity algorithms ^[31], others are adapted to ecological studies ^[32] ^[33] ^[34] and animal communication ^[35], and others are used in economics, popularized by the "Index of Economic Complexity' ^[36]. Similar indices have been developed in other Social Sciences and Law Studies ^[37]. For example, empirical work based on such indices showed that more complex constitutions may be harmful to society ^[38]. Ecologists have developed analytics for complexity measures systematically for a long time ^[39] ^[40] ^[41] ^[42]. Several attempts to relate complexity in thermodynamic terms have been published ^[43] ^[44] ^[45] ^[46] ^[47].

Multidimensional Complexity: Shannon information, and algorithmic information content can be combined to produce mathematical definitions for complexity that adapt to more complex systems by relating complexity with information [48]. Tackling high-dimensional problems with multidimensional vectors is common practice [49]. However, analytics of high complexity problems is better at explaining phenomena already known than predicting new ones [50] It has a mixed success in understanding highly complex systems. Even for low dimensional topology (4 or less dimensions) pure analytic methods have not been able to solve many issues [51]. At ever larger degrees of freedom, computational mathematical tools do not help in explaining phenomena such as Emergence [52]. Novel dimensions require new kinds of information and new metrics for its study. For example aggregates of atoms form molecules which may form cells that can organize into "organisms" that evolve brains: brains have properties and forms of storing and managing information that atoms do not have. Infodynamics [53] aims to understand complexity in multidimensional systems as an expression of information capable of producing new emergent properties in a given system. Classical thermodynamics refers to Free Energy as the energy that produces useful work, in contrast to Thermodynamic Entropy which refers to the energy that dissipates as heat and does not produce work. Analogously, the amount of free energy produced by a given amount of this information may serve as a way to understand complexity. This approach, called Infodynamics, has been tested empirically in over a dozen studies [54]. These included: Social complexity and colony size in ants increases as energy consumption per capita decreases; Economic development of countries increases as their scientific development expands; Per capita electricity consumption decreases in cities as their size increases; Countries with a strong Rule of Law have low infant mortality and a high Human Development Index; Countries with many populist words in their constitution underperformed in Human Development relative to those with simple constitutions.

Infodynamics

Infodynamics applies the logic of thermodynamics used in understanding the dynamics of energy, but applies it to



understand information dynamics ^[55]. It distinguishes between Thermodynamic and Information Entropy: The later relates to uncertainty in outcomes, while thermodynamic entropy pertains to energy distribution in physical systems. But the production of Thermodynamic Free energy is related to information ^[56].

Let's define Information Complexity I as the total amount of information in a system, and Useful Information as the one producing Free Energy F and thus work. Free Energy and Work are thermodynamic concepts so that

$$F = E - S$$

Where E is total energy and S the thermodynamic entropy due to energetic processes

If Φ is useful information or the information that accounts for F, I the total information accounting for its complexity, and N useless information or noise, then

$$\Phi = I - N$$
 and $F = E - S$ then $\Phi = k(E - S)$

were k is a function or constant relating \mathbf{F} with $\mathbf{\Phi}$

Here we have a tool to measure Φ quantitatively and empirically, where Infodynamic Complexity (Total Information I) can be measured as

$$I = \Phi + N + R$$

where R is redundant information which is important for conserving information in time.

This result relates the amount of information to the amount of work that can be produced by a system. This approach allows handling complex systems, including living organisms and ecosystems, and might be appropriate to tackle problems of quantification of information and useful work in Artificial Intelligence.

The quirk in complex irreversible processes is to differentiate between different kinds of information and the means to assess them. Separating "Useful Information" that produces "Free Energy" available to produce "Useful Work" from other kinds, such as redundant, useless information or noise is not a trivial exercise. The cost of information is only calculable indirectly. It might relate to the cost of engraving a substrate, reading it, storing it, communicating it, transmitting it through a medium, or the cost of acquiring it. For example, complexity of Large Language models in Artificial Intelligence, and thus of its information content, is often measured using the size of the databases used to train the models, or the size of the stored information. This method does not distinguish between noise, redundancy and useless data. Redundancy is relatively easy to measure by comparing the encrypted information with itself. Separating noise from useful information is possible by measuring its effect on the production of work. Large Language models can then be compared by their efficiency in producing useful information rather than by the energy expended in creating them, or by the size of the encrypted information they contain.

Discussion



Classical Physics works with only four fundamental physical dimensions. Acknowledging Godel's insight^[57], it is impossible to build a 4-d model of all atoms in the world with a human brain nor with any human build contraption. Thus, features that emerge from the interactions of systems more complex than those grasped by 4-d models, such as values, fitness, power and synergy are better treated as novel dimensions than as outcome of the 4-d interactions of zillions of subatomic particles. The task of the science of complexity is to define relevant new dimensions. One such is complexity and information.

Complexity and Information are closely linked. Complexity may be defined as the amount of knowledge that is required to describe a system. More complex systems require more information to describe. It also refers to patterns and representations. For example, a simple system can be described with a few simple properties, such as its size, shape, and color. A more complex system requires much more information. Describing a human requires describing its physical features, its personality, its memories, motivations, its thoughts, etc. This is a multidimensional complexity

In biological evolution natural selection forces genomes to behave as a natural "Maxwell Demon," within a fixed environment, genomic complexity is forced to increase [58]. Yet not all increase in complexity leads to an increase in useful information as shown above with the example of constitutions. Nor do organisms that have longer DNA chains in their genome have always a higher complexity than others with less DNA. The Australian lungfish has a genome 14 times larger than the human genome! That means that much information might be redundant or useless noise.

Many problems relating complexity with Infodynamics remain unresolved. For example:

- Crystals in inorganic chemistry are believed to minimize thermodynamic entropy and information entropy of the system.
 Non-symmetric heterogeneous crystals such as DNA, RNA or protein crystals do not. We lack descriptive tools to better understand these differences.
- The depth of resolution of details of a system influences its metrics of information and complexity. It is not the same to
 count organisms in a system than to count cells, atoms or subatomic particles in the same system. Solving a problem
 requires the appropriate resolution of details of the system. Defining the appropriate resolution for studying specific
 problems is a challenge.
- Structural Information is always present and not evident when exploring a problem that does not include the dimensions defining the border condition of the system appropriately. For example, what are the structural features that can be considered as relevant information when analyzing the free energy gained by the total energy released by an explosive inside a cannon? What are the relevant processes that allow algorithms with better intelligence? Trial and error is often the only way to figure this out, making empirical studies indispensable.

When equating Complexity with Information, we have to keep in mind that different types of information exist and thus different kinds of complexity. Information can be structural, enformation, intropy, entangled, encrypted, redundant, synergic or noise, as recognized by Infodyamics ^[59]. The most important type of information is what Infodynamics calls useful information Φ : the one that allows systems to produce Free Energy that produces Useful Work. An outline for a taxonomy of complexity that ranges from simple complexity measured in bits, to evolutionary products such as human



brains, or social dynamic complexity such as modern science, is proposed here. Tackling measurements of complexity empirical, such as engineers do, is made possible by Infodynamics, and increases our understanding of the world of complexity.

Understanding the basics of thermodynamics equips engineers with the necessary tools to analyze energy systems and enhance their efficiency. It provides a framework for solving complex problems in energy management, system optimization, and sustainability, which are central concerns in today's engineering challenges. Analogously, understanding the basic working of information and its relation to the production of free thermodynamic, biological or cultural energy, will allow us to better understand general intelligence and emergence of new knowledge. This understanding is at the root of any progress in complex systems science and emergent intelligences.

Having measures of complexity that allow different systems to be compared by applying a common metric is fundamental in stabilizing the foundations of Infodynamics. Measures of complexity allow different systems to be compared to each other by applying a common metric. This is especially meaningful for systems that are structurally or functionally related. Differences in complexity among such related systems may reveal features of their organization that increase our understanding of complexity and information.

References

- 1. Peliti, L., & Vulpiani, A. (Eds.). (1988). Measures of complexity. Berlin, Germany:: Springer.
- 2. ^Lopez-Ruiz, R., Mancini, H. L., & Calbet, X. (1995). A statistical measure of complexity. Physics letters A, 209(5-6), 321-326.
- 3. Vovk, V., Papadopoulos, H., & Gammerman, A. (2015). Measures of Complexity. Springer.
- 4. Wiesner, K., & Ladyman, J. (2019). Measuring complexity. arXiv preprint arXiv:1909.13243.
- 5. ^Heylighen, Francis (1999). The Growth of Structural and Functional Complexity during Evolution, in; F. Heylighen, J. Bollen & A. Riegler (Eds.) The Evolution of Complexity. (Kluwer Academic, Dordrecht): 17-44.
- 6. ^Simplifying complexity: a review of complexity theory. Steven M Manson. Geoforum Volume 32, Issue 3, August 2001, Pages 405-414. https://doi.org/10.1016/S0016-7185(00)00035
- 7. ^Rosen, R. (1986). On information and complexity. In Complexity, language, and life: Mathematical approaches (pp. 174-196). Berlin, Heidelberg: Springer Berlin Heidelberg.
- 8. Traub, J. F., & Werschulz, A. G. (1998). Complexity and information (Vol. 26862). Cambridge University Press.
- 9. Traub, J. F. (2003). Information-based complexity. In Encyclopedia of Computer Science (pp. 850-854).
- 10. ^Byström, K. (1999). Task complexity, information types and information sources: examination of relationships. Tampere University Press.
- 11. ^Kolmogorov, Andrey (1963). "On Tables of Random Numbers". Sankhyā Ser. A. 25: 369-375. MR 017848
- 12. *Fisher, R. A. (1922-01-01). "On the mathematical foundations of theoretical statistics". Philosophical Transactions of the Royal Society of London, Series A. 222 (594-604): 309-368. doi:10.1098/rsta.1922.0009. hdl:2440/15172.
- 13. ^Shannon, C. (1948) A mathematical theory of communication. Bell System Technical Journal 27 379-423.



- 14. ^Johnson, Neil F. (2009). "Chapter 1: Two's company, three is complexity" (PDF). Simply complexity: A clear guide to complexity theory. Oneworld Publications. p. 3. ISBN.
- 15. ^Heylighen, F., Cilliers P., Gershenson C. Complexity and Philosophy. arXiv:cs/0604072
- 16. Slocombe L., Wlker S.I. Measuring Molecular Complexity. ACS Cent. Sci 2024 doi/10.1021/acscentsci.4c00697
- 17. Nasios, I. (2024). Analyze mass spectrometry data with artificial intelligence to assist the understanding of past habitability of Mars and provide insights for future missions. Icarus, 408, 115824.
- 18. ^Whitesides G.M., Ismagilov R.F. Complexity in Chemistry. SCIENCE, 1999, 284, Issue 5411, pp. 89-92. DOI: 10.1126/science.284.5411.89
- 19. ^Chakraborty, S., Vinodchandran, N. V., & Meel, K. S. (2023). Distinct Elements in Streams: An Algorithm for the (Text) Book. arXiv preprint arXiv:2301.10191.
- 20. Feutrill, A., & Roughan, M. (2021). A review of Shannon and differential entropy rate estimation. Entropy, 23(8), 1046.
- 21. ^Febres, G., & Jaffe, K. (2017). Music viewed by its entropy content: A novel window for comparative analysis. PloS One, 12(10), e0185757.
- 22. ^Febres, G., & Jaffé, K. (2016). Calculating entropy at different scales among diverse communication systems. Complexity, 21(S1), 330-353.
- 23. Orlov, Y. L., Te Boekhorst, R., & Abnizova, I. I. (2006). Statistical measures of the structure of genomic sequences: entropy, complexity, and position information. Journal of bioinformatics and computational biology, 4(02), 523-536.
- 24. ^Davis, J. S., & LeBlanc, R. J. (1988). A study of the applicability of complexity measures. IEEE transactions on Software Engineering, 14(9), 1366-1372.
- 25. Li, M., & Vitányi, P. (1988). Two decades of applied Kolmogorov complexity.
- 26. ^Shiner, J. S., Davison, M., & Landsberg, P. T. (1999). Simple measure for complexity. Physical Review E, 59(2), 1459.
- 27. ^Lloyd, S. (2001). Measures of complexity: a nonexhaustive list. IEEE Control Systems Magazine, 21(4), 7-8.
- 28. Levy-Carciente, S., Sabelli, H., & Jaffe, K. (2004). Complex patterns in the oil market. Interciencia, 29(6), 320-323.
- 29. ^Mantegna, R. N., & Stanley, H. E. (1999). Introduction to econophysics: correlations and complexity in finance.

 Cambridge university press.
- 30. Zenil, H., Kiani, N. A., & Tegnér, J. (2018). A review of graph and network complexity from an algorithmic information perspective. Entropy, 20(8), 551.
- 31. ^Ball, M. O. (1986). Computational complexity of network reliability analysis: An overview. leee transactions on reliability, 35(3), 230-239.
- 32. ^Landi, P., Minoarivelo, H.O., Brännström, Å., Hui, C. and Dieckmann, U. (2018), Complexity and stability of ecological networks: a review of the theory. Popul Ecol, 60: 319-345. https://doi.org/10.1007/s10144-018-0628-3
- 33. ^Measuring ecological complexity. Lael Parrott. Ecological Indicators. Volume 10, Issue 6, November 2010, Pages 1069-1076
- 34. ^Landi, P., Minoarivelo, H. O., Brännström, Å., Hui, C., & Dieckmann, U. (2018). Complexity and stability of ecological networks: a review of the theory. Population ecology, 60(4), 319-345.
- 35. ^IACOPINI, lacopo, et al. Not your private tête-à-tête: leveraging the power of higher-order networks to study animal



- communication. Philosophical Transactions B, 2024, vol. 379, no 1905, p. 20230190.
- 36. ^Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the national academy of sciences, 106(26), 10570-10575.
- 37. Ma, J., Wen, G., Wang, C., & Jiang, L. (2019). Complexity perception classification method for tongue constitution recognition. Artificial intelligence in medicine, 96, 123-133.
- 38. ^Jaffe, K., Contreras, J. G., Soares, A. C., Correa, J. C., Martinez, E., & Canova, A. (2021). The Relationship between Constitutions, Socioeconomics, and the Rule of Law: A Quantitative Thermodynamic Approach. Socioeconomics, and the Rule of Law: A Quantitative Thermodynamic Approach (August 3, 2021).
- 39. Parrott, L. (2010). Measuring ecological complexity. Ecological Indicators, 10(6), 1069-1076.
- 40. ^Toussaint, O., & Schneider, E. D. (1998). The thermodynamics and evolution of complexity in biological systems.

 Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 120(1), 3-9.
- 41. Loehle, C. (2004). Challenges of ecological complexity. Ecological complexity, 1(1), 3-6.
- 42. ^Bascompte, J., & Solé, R. V. (1995). Rethinking complexity: modeling spatiotemporal dynamics in ecology. Trends in Ecology & Evolution, 10(9), 361-366.
- 43. Fivaz, R. (1991). Thermodynamics of complexity. Systems Research, 8(1), 19-32.
- 44. ^Mikulecky, D. C. (2001). Network thermodynamics and complexity: a transition to relational systems theory. Computers & chemistry, 25(4), 369-391.
- 45. Lloyd, S., & Pagels, H. (1988). Complexity as thermodynamic depth. Annals of physics, 188(1), 186-213.
- 46. ^Lizier, J. T. (2014). JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. Frontiers in Robotics and AI, 1, 11.
- 47. Salthe, S. N. (2001). What is Infodynamics?. In Understanding complexity (pp. 31-38). Boston, MA: Springer US.
- 48. Gell-Mann, Murray, and Seth Lloyd. "Information measures, effective complexity, and total information." Complexity 2, no. 1 (1996): 44-52.
- 49. ^Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. Cognitive computation, 1, 139-159.
- 50. [^]Berglund, P., Hübsch, T., & Minic, D. (2023). On de Sitter spacetime and string theory. International Journal of Modern Physics D, 32(09), 2330002.
- 51. ^Harnett K. (2024)A New Agenda for Low-Dimensional Topology www.quantamagazine.org/a-new-agenda-for-low-dimensional-topology-20240222/
- 52. ^Jaffe K. (2023). Thermodynamics, Infodynamics and Emergence. Qeios. doi:10.32388/S90ADN.6
- 53. ^Jaffe K. (2024). Infodynamics, a Review. Qeios. doi:10.32388/2RBRWN.4
- 54. ^Jaffe K. (2023). A Law for Irreversible Thermodynamics? Synergy Increases Free Energy by Decreasing Entropy.

 Qeios. doi:10.32388/2VWCJG.5
- 55. ^Jaffe K (2024) Infodynamics, Information Entropy and the Second Law of Thermodynamics. Qeios T13JP9.3 eios.com/read/T13JP9.3
- 56. Maxwell J. C. Theory of Heat; Longmans, Green, and Co.: London, UK, 1871; Chapter 12, (1871)
- 57. ^Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.



Monatshefte für mathematik und physik, 38, 173-198.

- 58. ^Adami, C., Ofria, C., & Collier, T. C. (2000). Evolution of biological complexity. Proceedings of the National Academy of Sciences, 97(9), 4463-4468
- 59. ^Jaffe K. (2024). Infodynamics, a Review. Qeios. doi:10.32388/2RBRWN.4