Article

Entropy and Wealth

Demetris Koutsoyiannis* and G.-Fivos Sargentis

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, 15780 Athens, Greece

* Correspondence: dk@itia.ntua.gr

Abstract: While entropy was introduced in the second half of the 19th century in the international vocabulary as a scientific term, in the 20th century it became common in colloquial use. Popular imagination has loaded "entropy" with almost every negative quality in the universe, in life and in society, with a dominant meaning of disorder and disorganization. Exploring the history of the term and many different approaches on it, we show that entropy has a universal stochastic definition which is not disorder. The accompanying principle of maximum entropy, which lies behind the Second Law, gives explanatory and inferential power to the concept and promotes entropy as the mother of creativity and evolution. As social sciences are often contaminated by subjectivity and ideological influences, we try to explore whether the maximum entropy, applied to the distribution of wealth quantified by annual income, can give an objective description. Using publicly available income data, we show that the income distribution is consistent with the principle of the maximum entropy. The increase of entropy is associated to increase of society's wealth yet a standardized form of entropy can be used to quantify inequality. Historically, technology has played a major role in development and increase of the entropy of income. Such findings are contrary to the theories of ecological economics and other theories which use the term entropy in a Malthusian perspective.

Keywords: entropy; wealth; income distribution; options; potentiality; principle of maximum entropy; Second Law; stochastics

Wealth is not about having a lot of money; it's about having a lot of options
Chris Rock (American comedian, writer, producer and director)

Πάντα τὰ ἐμὰ μετ᾽ ἐμοῦ φέρω (All that is mine I carry with me)
Bias of Priene (one of the seven Greek sages; 6th c. BC;
quoted in Latin by Cicero, Paradoxa Stoicorum I, 8, as "Omnia mea mecum porto")

1. Introduction

The name "entropy" has been introduced about 150 years ago as a scientific term but later its use became common of everyday language. We can find it in literature [1,2] in poetry [3,4], in press, and in web posts, but often its use is irrelevant to its real scientific meaning. The most common use of the word entropy is when a writer wants to describe with an "intellectual" word a kind of disorder. We will clarify in detail in section 2.3 that this is a misinterpretation of the actual meaning of the term, which in fact is more related to uncertainty. The wide colloquial use of entropy becomes clear if we see the results from the base of detailed information on the top 60 000 words (lemmas) of English, based on data from the Corpus of Contemporary American English (COCA), whose content includes eight genres: spoken, fiction, popular magazines, newspapers, academic texts and more [5]. According to data given by Word and Phrase Info [6] and plotted in Figure 1, the frequency of "entropy" in fiction, for example, is not dramatically lower than in the academic texts.

The academic corpus can be investigated using bibliometric databases, such as Scopus [7]. The results obtained from several searches on the latter are shown in Figure 2. One would expect that the term "entropy" would more frequently appear in scholar articles in combination with terms such as "physics" or "thermodynamics"—and indeed this is the case for the recent years. Amazingly however, before 1960s the combination of term "entropy" with "society" or "social" was more frequent than the former. This suggests the appeal of a concept related to the Second Law of thermodynamics in social sciences. We note that in that period the probabilistic content of entropy (see section 2) was not fully developed and, thus, in papers before the 1960s, the entropy was used with the classical thermodynamical meaning (also explained in section 2). Specifically, the resource flow in economics was parallelized with the energy flow in thermodynamics. As seen in the figure, in the 21st century "entropy" is also used in combination with ecology and economics.

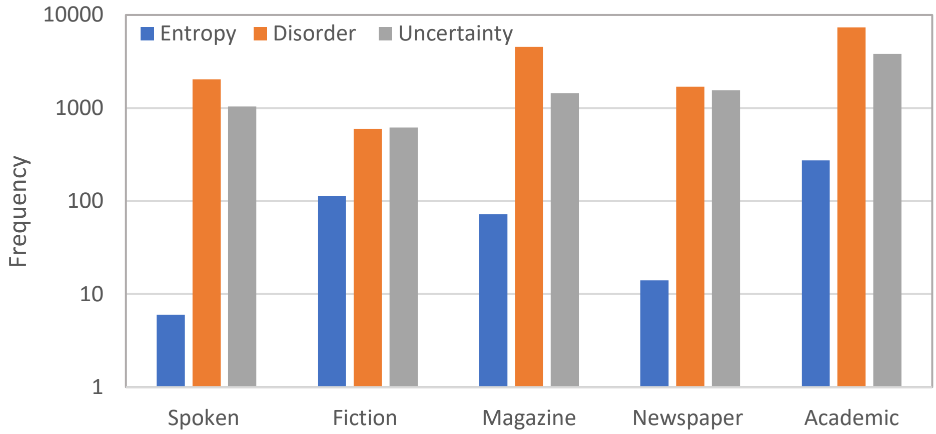


Figure 1. Frequency of appearances of the words: entropy, disorder and uncertainty in the Corpus of Contemporary American English. Data from [5,6].

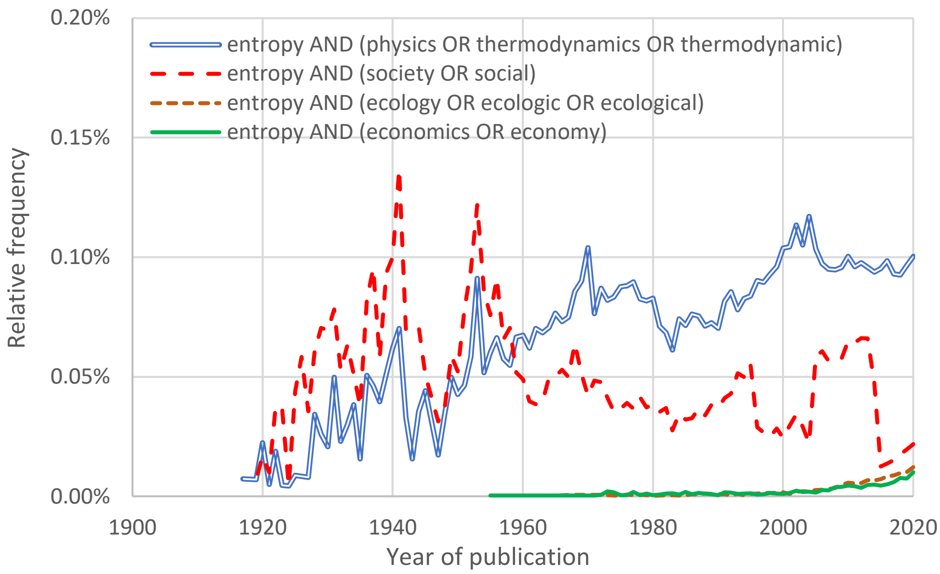


Figure 2. Relative frequency of appearances of the indicated key phrases in the article title, abstract and keywords of about 70 million articles written in English, which are contained in the Scopus database [7] up to year 2020.

Out of its physical and stochastic context, the term "entropy" is typically used metaphorically and hence its meaning becomes ambiguous or diverse. For example, the term "social entropy", in one of its earliest uses in scholarly publications [8] is equated to "dereliction, pollution and waste", which are created by "economic activity" or by "society as consumers" and has to be minimized. Bailey in his book entitled "Social Entropy Theory" [9] tried to illuminate *the fundamental problems of societal analysis with a nonequilibrium approach, a new frame of reference built upon contemporary macrological principles, including general systems theory and information theory.*

His interest is more to illuminate the "Social Entropy Theory" than define social entropy per se. Nor in an overview of his book [10], did Bailey provide a definition of social entropy. In a critique of the book, Mayer [11] finds the "unrelenting abstractness of Social Entropy Theory quite frustrating" and adds:

Never is the theory applied to real sociological data or anything like a real social situation.

Neto et al. [12], who also use Bailey' theory, provide little help in clarifying what social entropy is. Recently, Dinga et al. [13] building on Bailey's theory, clarified, rather qualitatively, the concept of social entropy as follows: (i) it cannot be of a Thermodynamic type; (ii) it must be connected with social order; (iii) its connection with social order must be inversely proportional; (iv) it must hold the connotations of the relationship between the homogeneity and heterogeneity of a system/process; and (v) it is fundamentally grounded by normativity. Interestingly however, Balch [14] used the term "social entropy" as a measure of robot group diversity, proposing a formal mathematical definition.

Different aspects of entropy and energy in social systems have been examined by Mavrofides et al. [15], as well as by Davis [16], who examines how the physical and mathematical notions of entropy can be usefully imported into the social sphere. Davis also uses the term "social entropy" without a definition. Further, he states:

Entropy has been characterized variously and vaguely by the words decadence, decay, anarchy, dystopia, randomness, chaos, decay, waste, inefficiency, dissipation, loss of available energy, irreversibility, indeterminacy, most probable states, equilibrium, thermal death […]. *In the social sphere it has been characterized as apocalypse, disorder, disorganization, disappearance of distinctions, meaninglessness, absurdity, uncertainty, pandemonium, loss of information, inert uniformity, incoherence.* […] *In* [humanistic] *areas, the concept is used more or less as a metaphor or a synonym for chaos, disorder, breakdowns, dysfunctions, waste of material and energy, enervation, friction, inefficiencies.*

Davis is influenced by Saridis [17], according to whom "entropy measures the waste produced when work is done for the improvement of the quality of human life"; he highlights the following quotation by Saridis:

The concept of Entropy creates a pessimistic view for the future of our universe. The equalization of all kinds of sources of activities is leading to the equivalent of thermal death and universal boredom of our world.

In economics, Frederick Soddy (1877 –1956) [18-20] and Nicholas Georgescu-Roegen (1906 –1994) [21-23], fascinated by entropy in thermodynamics, sought analogies with economics and development in a Malthusian perspective. Avery [24] notes that:

Early in the 20th century, both Frederick Soddy and Nicholas Georgescu-Roegen discussed the relationship between entropy and economics. Soddy called for an index system to regulate the money supply and a reform of the fractional reserve banking system, while Georgescu-Roegen pointed to the need for Ecological Economics, a steady-state economy, and population stabilization.

Following them, a series of papers and books studied similarities between economics and thermodynamical entropy [21,25-38]. McMahon and Mrozek used entropy, within the context of neoclassical economic thought, as a limit to economic growth [39]. In the same spirit, Smith and Smith used the Second Law of thermodynamics also to determine limits to growth [40]. In a review paper, Hammond and Winnett [41], present the influence of thermodynamics on the emerging transdisciplinary field of ecological economics. However, Kovalev [42] claims that entropy cannot be used as a measure of economic scarcity.

In a different context, in their editorial note in a special issue on “Maximum Entropy Economics: Foundations and Applications”, Scharfenaker and Yang [43] ask: “Maximum entropy economics: where do we stand?”. In reply, the same authors [44] offer a brief overview of what they consider the state of maximum entropy reasoning in economic research. These use a probabilistic (or information based) definition of entropy and so does Ryu [45] who presents a technique to determine the functional forms of income distributions maximizing entropy under given conditions. Fu et al. [46] use entropy divergence methods to define measures of income inequality; notably, they regard the uniform distribution of income the one with “the least inequality” (see discussion on this in section 4.1 below). Mayer et al. [47] provide a theoretical framework for studying the dynamics of coupled human and natural systems in an attempt to define sustainability.

Apparently, the above overview of entropy in social sciences and particularly in economics is not complete. Our purpose is not to review all related works but to highlight two facts. First, that the use of the notion of entropy is mainly metaphorical, rich in imaginary interpretation and divergent. And, second, that the dominant view is that entropy epitomizes all “bad things” one can think in the universe, in life, in human societies and in economics.

Our own view is quite different. On the first issue we insist that entropy should be used as a rigorous mathematical (in particular, stochastic) concept. We avoid using ambiguous terms such as “social entropy”. We claim that any interpretation should be as close to the mathematical definition as possible, and free of metaphoric meanings. On the second issue, we believe that the overloading of the concept of entropy with negative properties reflects misunderstanding of the underlying theory, guided by a deterministic world view—in which however entropy has no place.

We clarify our own view of entropy and its meaning in section 2, after tracing its roots as a scientific concept and its historical evolution in the last 150 years. In section 3 we provide a formal presentation of the principle of maximum entropy and its results under conditions relevant to material wealth. In section 4 we apply the framework on the economy, trying to show that the principle of maximum entropy explains the general behaviours seen in economics.

We try to provide several theoretical and even philosophical insights on important issues on entropy and economy, which unavoidable are influenced by our own perception—clearly an optimistic one, contrary to the pessimism expressed in most of papers reviewed above. At the same time, we try to infer these insights based on real world data, rather than speculation. The data we use are freely available on the internet and the reader can retrieve them and reproduce our calculations, or check and reprocess them independently.

Finally, we try to make the paper self-contained and stand-alone, so that even a reader unfamiliar with entropy, with only basic knowledge on calculus and probability, could assimilate it. The mathematical framework we develop can readily be put to work on the simplest computational framework (e.g. a spreadsheet).

2. What is entropy?

2.1 The origin of the entropy concept

The name *ἐντροπία* (Greek for entropy) appears already in ancient Greek [48] (from the verb ἐντρέπειν, to turn into, to turn about) but was introduced in the international scientific vocabulary by Rudolf Clausius only in 1865 (although the concept appears also in his earlier works, as he described in [49]). The rationale for introducing the term is explained in his own words [50, p. 358], which indicate that he was not aware of the existence of the very word in ancient Greek:

> *We might call S the transformational content of the body […]. But as I hold it to be better to borrow terms for important magnitudes from the ancient languages, so that they may be adopted unchanged in all modern languages, I propose to call the magnitude S the entropy of the body, from the Greek word τροπή, transformation. I have intentionally formed the word*

entropy so as to be as similar as possible to the word energy; for the two magnitudes to be denoted by these words are so nearly allied their physical meanings, that a certain similarity in designation appears to be desirable.

In addition to its semantic content, this quotation contains a very important insight: the recognition that entropy is related to transformation and change, and the contrast between entropy and energy, where the latter is a quantity that is conserved in all changes. This meaning has been more clearly expressed in Clausius' famous aphorism **[**51**]**:

Die Energie der Welt ist constant. Die Entropie der Welt strebt einem Maximum zu.
(The energy of the world is constant. The entropy of the world strives to a maximum).

In other words, entropy and its ability to increase (as contrasted to energy, momentum and other quantities that are conserved) is the driving force of change. This property of entropy is acknowledged only seldom [52-54]. Instead, as we have already seen, in common perception entropy epitomizes all "bad things" as if it were disconnected with change, or as if change can only have negative consequences, always leading to deterioration.

Mathematically, the thermodynamic entropy, S, is defined in the same Clausius' texts through the equation $\mathrm{d}S = \delta Q/T$, where Q and T denote heat and temperature. The definition, however, applies to a reversible process only. The fact that in an irreversible process $\mathrm{d}S > \delta Q/T$ makes the definition imperfect and affected by circular reasoning, as, in turn, a reversible process is one in which the equation holds.

Two decades later (in 1877) Ludwig Boltzmann [55] (see also Swendsen, [56]) gave entropy a probabilistic content as he linked it to probabilities of statistical mechanical system states, thus explaining the Second Law of thermodynamics as the tendency of the system to run toward more probable states, which have higher entropy. The probabilistic concept of entropy was advanced later in thermodynamics by the works of Gibbs [57].

The next important step was made by Shannon in 1948 [58]. Shannon used an essentially similar, albeit more general, entropy definition to describe the information content, which he also called entropy at von Neumann's suggestion [59-61]. According to the latter definition, entropy is a probabilistic concept, a measure of information or, equivalently, uncertainty. In the same year, Wiener, in his famous book *Cybernetics*, also published in 1948 [62], used the same definition for information, albeit with a negative sign (p. 62) because he regarded the information as the negative of entropy (p. 11). (Interestingly, he formed the celebrated term *Cybernetics* from the Greek word *κυβερνήτης*, meaning steersman, pilot, skipper, governor, albeit incorrectly spelling it in his book —p. 11— as *χυβερνήτης*.)

A few years later, in 1956, von Neumann [63] obtained virtually the same definition of entropy as Shannon, in a slightly different manner. Notably, as von Neumann, in addition to being a mathematician and computer scientist, was also a physicist, engineer and polymath, he clearly understood the connection of the probabilistic definition of entropy with its pre-existing physical content. Specifically, he wrote:

An important observation about this definition is that it bears close resemblance to the statistical definition of the entropy of a thermodynamical system. […] Pursuing this, one can construct a mathematical theory of the communication of information patterned after statistical mechanics.

He also cited an earlier work (1929) in physics by Szilard [64], who implied the same definition of entropy in a thermodynamic system.

The last fundamental contribution of the entropy concept was made a year later (in 1957) by Jaynes [65], who introduced the *principle of maximum entropy*. This postulates that the entropy of a stochastic system should be at maximum, under some conditions, formulated as constraints, which incorporate the information that is given about this system. This principle can be used for logical inference as well as for modelling physical systems. In this respect, the tendency of entropy to become maximal (Second Law of thermodynamics), which drives natural change, can result from this principle. On the other hand, the principle equips the entropy concept with a powerful tool for logical inference.

2.2 Are thermodynamic and probabilistic entropy different?

More than 150 years after the introduction of the entropy concept, its meaning is still debated, and a diversity of opinion among experts is encountered [66]. In particular, despite having the same name, the probabilistic (or information) entropy and the thermodynamic entropy are still regarded by many (perhaps the majority of scientists) as two distinct notions having in common only the name. The classical definition of thermodynamic entropy (as above) does not give any hint about similarity with the probabilistic entropy. The fact that the latter is a dimensionless quantity and the former has units (J/K) has been regarded as an argument that the two are dissimilar. Even Jaynes (2003), the founder of the maximum entropy principle, states:

> *They should never have been called by the same name; the experimental entropy makes no reference to any probability distribution, and the information entropy makes no reference to thermodynamics. Many textbooks and research papers are flawed fatally by the author's failure to distinguish between these entirely different things.*

However, the units of thermodynamic entropy are only an historical accident, related to the arbitrary introduction of temperature scales [54]. Furthermore, the connection of probabilistic and thermodynamic entropy is clearly implied by its pioneers, Boltzmann [55], Gibbs [57], Szilard [64] and von Neumann [63]. A more recent account of the connection has been provided by Robertson [59]. Furthermore, as has recently been shown [67,68], the thermodynamic entropy of gases can be easily produced by the formal probability theory without the need of strange assumptions (e.g. indistinguishability of particles). Impressive examples of deductive reasoning in deriving thermodynamic laws from the formal probabilistic principle of maximum entropy have been provided in [68]. Notable among them is the derivation of the law of phase transition of water (Clausius-Clapeyron equation) by maximizing entropy, i.e. uncertainty, at the microscopic level, of a single water molecule, yet leading to an expression that is virtually certain at a macroscopic level.

2.3 Does entropy measure disorder?

As already mentioned, in the public perception entropy has a negative content, and is typically identified with disorganization or disorder, and deterioration. This misleading perception has its roots in the scientific community, albeit not in the founders of the concept (except one, as we will see). Boltzmann did not identify entropy with disorder, even though he used the latter word in a footnote appearing in two papers of his [69,70], in which he speaks about the

> *agreement of the concept of entropy with the mathematical expression of the probability or disorder of a motion.*

Clearly, he speaks about the irregular motion of molecules in the kinetic theory of gases for which his expression makes perfect sense. Boltzmann also used the notion of disorder with the same meaning, in his Lectures on Gas Theory [71]. On the other hand, Gibbs [57], Shannon [58] and von Neumann [63] did not use the term disorder or disorganization at all.

One of the earliest (in 1944) uses of the term disorder is in a paper by Darrow [72] in which he states:

> *The purpose of this article has been to establish a connection between the subtle and difficult notion of entropy and the more familiar concept of disorder. Entropy is a measure of disorder, or more succinctly yet, entropy is disorder: that is what a physicist would like to say.*

Epistemologically, it is interesting that a physicist prefers the "more familiar" but fuzzy concept of disorder over the "subtle and difficult", yet well-defined at his time, concept of entropy.

However, it appears that it was Wiener the most influential scientist who supported the disorder interpretation. In 1946, he gave a keynote speech at the New York Academy of Sciences [73] in which he declared that:

> *Information measures order and entropy measures disorder.*

Also, in his influential book Cybernetics [62, p. 11], he stated:

> *the entropy of a system is a measure of its degree of disorganization*

where he replaced the term disorder with disorganization, as in this book he extensively used the former term for mental illness.

Even in the 21st century, the disorder interpretation is dominant. For example, Chaitin [74] stated:

> *Entropy measures the degree of disorder, chaos, randomness, in a physical system. A crystal has low entropy, and a gas (say, at room temperature) has high entropy.*

More recently, Bailey [75] claimed:

> *As a preliminary definition, entropy can be described as the degree of disorder or uncertainty in a system. If the degree of disorder is too great (entropy is high), then the system lacks sustainability. If entropy is low, sustainability is easier. If entropy is increasing, future sustainability is threatened.*

It is relevant to remark in the latter quotations that disorder has been used as equivalent to uncertainty or randomness—where the latter two terms are in essence identical [76]. Furthermore, the claim that a high entropy system lacks sustainability is at least puzzling, given that the highest entropy occurs when a system is at the most probable (and hence most stable) state.

There is no doubt that the notion of entropy entails difficulties in understanding but this happens because our education is based on the deterministic paradigm. Indeed, it is difficult to incorporate a clearly stochastic concept, i.e. entropy, in a deterministic mindset. The notion of order looks deterministic-friendly, and its opposite, disorder, has a negative connotation in the deterministic mindset.

But the notions of order and disorder are less appropriate and less rigorous as scientific terms and more appropriate in describing mental states (as in Wiener's use described above; cf. personality disorder, stress disorder, bipolar disorder, mental disorder), and even more so in describing socio-political states. The latter is manifest in the frequent use of the expression "world order" and "new world order" in political texts included in Google Books (see Figure 3). As a further example, the phrases "world order" and "new world order" were the favourite ones of Kissinger, the American geopolitical consultant who served as United States Secretary of State and National Security Advisor. Specifically, they appear in 109 and 28 of his articles registered in Google Scholar, respectively [77,78], of which an article of 2009 [79] was up to recently his most popular (48 citations). However, the variant "new world order" used by Kissinger in [79] has become infamous, as had the Nazis' "new order" after the World War II (see Figure 3). Naturally, Kissinger changed it in his article of 2020 [80], which by now has become even more popular (132 citations) to "liberal world order", using it for first time but borrowing it from other authors (e.g. [81]). As will be explained in the next few lines, the latter expression is self-contradictory or a euphemism.

Other representatives of oligarchic elites prefer the expression "global order" (also included in Figure 3). For example, the recent World Economic Forum's book "Covid-19: The Great Reset" [82], the latter expression appears seven times (and once the "global disorder"), while "new order" and "world order" do not appear at all (except in a reference to Kissinger's article). Note that the book invokes the "Covid-19 pandemic" (appearing 14 times) along with "climate change" (appearing 37 times), "global warming" (appearing four times), or "climate crisis" (appearing twice) to promote the idea of a "great reset" (appearing 13 times). The latter comprises economic reset, societal reset, geopolitical reset, environmental reset, industry and business reset, and even individuals reset.

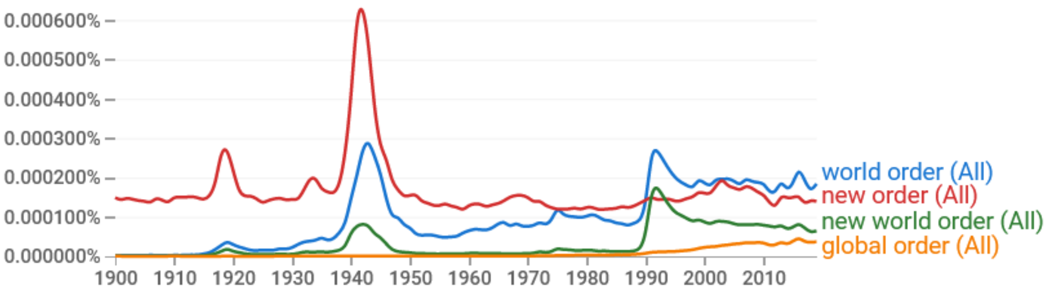


Figure 3. Frequency of appearances of the indicated phrases in Google Books [83,84]. Notice that "new order" was the political order which Nazi Germany wanted to impose and naturally its use peaked in the early 1940s. The other phrases also peaked at the same time but they showed higher peaks after 1990s.

In one of the earliest critiques of the disorder interpretation of entropy, Wright [85], makes a plea for moderation in the use of "intuitive qualitative ideas concerning disorder". At a more absolute tone, recently Leff [86] states:

The too commonly used disorder metaphor for entropy is roundly rejected.

In an even more recent article, Styer [87] states:

we cannot stop people from using the word "entropy" to mean "disorder" or "destruction" or "moral decay." But we can warn our students that this is not the meaning of the word "entropy" in physics.

Steyer attributes an excessive contribution to the misconception of entropy as disorder to autobiographical book "The Education of Henry Adams" [88]. He relates that it proved to be enormously influential, as it won the 1919 Pulitzer Prize in biography, and in April 1999 was named by Modern Library the 20th century's best nonfiction book in English. As quoted by Steyer, Adams contests Chaos and Anarchy, and states:

The kinetic theory of gas is an assertion of ultimate chaos. In plain words, Chaos was the law of nature; Order was the dream of man.

This looks to be a very strong statement. Undoubtedly, elites that want to control the world have exactly this dream (cf. [89] and references above). But this does not necessarily mean that the entire humanity has the same dream with the elites. When speaking about entropy, we should have in mind that the scale is an important element and that entropy per se, being a probabilistic concept, presupposes a macroscopic view of phenomena, rather than a focus on individuals or small subsets. If we viewed the motion of a particular dice throw, we might say that it was irregular, uncertain, unpredictable, chaotic, random. But macroscopization, by removing the details, may also remove irregularity. For example, the application of the principle of maximum entropy to the outcomes of a die results in equal probabilities (1/6) of each outcome. This is perfect order macroscopically. Likewise, as already mentioned, the maximum uncertainty in a particular water molecule's state (in terms of position, kinetic state and phase), macroscopically results in the Clausius-Clapeyron law. Again, we have perfect order, as the accuracy of this law is so high that most people believe that it is a deterministic law.

But if entropy is not disorder, what is it? This question is not that difficult to answer, as the above discourse implies. According to its standard definition, which will be repeated in section 2.6, entropy is precisely the expected value of the minus logarithm of probability. If this sounds too difficult to interpret, an easy and accurate interpretation (again explained in section 2.6) is that entropy is a measure of uncertainty. Hence, maximum entropy means the maximum uncertainty that is allowed in natural processes, given the constraints implied by natural laws.

If "disorder" is regarded as a "bad thing", for many the same is the case with uncertainty. The expressions "uncertainty monster" and "monster of uncertainty" appear in about 250 scholarly articles registered in Google Scholar (samples are [90,91], to mention a couple of the most cited with the word "monster" appearing in their title). However, if uncertainty is a monster, it is thanks to this monster that life is liveable and fascinating. Uncertainty is not an enemy of science or of life; rather it is the mother of creativity and evolution. Without uncertainty, life would be a "universal boredom" (to borrow a phrase by Saridis [17] and reverse its connotation), and concepts such as hope, will (particularly, free will), freedom, expectation, optimism, etc., would hardly make sense. A technocratic system where an elite comprising super-experts who, using super-models, could predict the future without uncertainty, would also assume full control on the society [92]. Fortunately, this will never happen because entropy, i.e. uncertainty, is a structural property of nature and life. Hence, in our view, uncertainty is neither disorder nor a "bad thing". How could the most important law of physics (the Second Law) be a "bad thing"?

In a deterministic world view, there is no uncertainty and there is no meaning in speaking about entropy. If there is no uncertainty, each outcome is accurately predicted and hence there are no options. In contrast, in an indeterministic world, there is a plurality of options. This corresponds to the Aristotelian idea of *δύναμις* (Latin: *potentia*—English: *potency* or *potentiality*). The existence of options entails that there is freedom, in the following sequence:

entropy ↔ uncertainty ↔ plurality of options ↔ freedom.

This view, also depicted in Figure 4, is consistent with what was vividly expressed by Brissaud [60]:

> *Entropy measures freedom, and this allows a coherent interpretation of entropy formulas and of experimental facts. To associate entropy and disorder implies defining order as absence of freedom.*

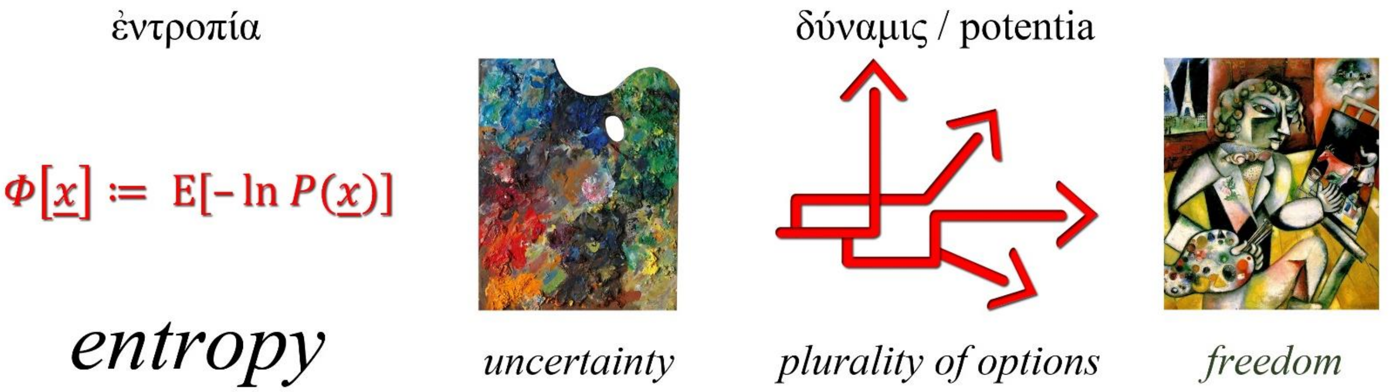


Figure 4. An attempt for an artistic representation of the notion of entropy. Uncertainty is depicted by Marc Chagall's Palette (adapted from [93]) and freedom by Marc Chagall's *Self-Portrait with Seven Fingers* [94]; *δύναμις* (Greek) or *potentia* (Latin) is the Aristotelian idea for potency or potentiality.

2.4 On negentropy

In 1921 the Swiss physicist, C.-E. Guye [95] (followed by other scientists) asked the question: How is it possible to understand life, when the whole world is ruled by such a law as the second principle of thermodynamics, which points toward death and annihilation? Today it makes sense to ask: Has this question been answered by now? Or, is it still relevant, one hundred years after? As insightfully discussed by Brillouin (1949 [96]), scientists of the era wondered if there was a "life principle", a new and unknown principle that would explain life as an entity that contrasts the second law of thermodynamics. A year after, Brillouin coined the term *negentropy* as an abbreviation of negative entropy [97]. In this, he used information theoretic concepts to express the idea that every observation in a laboratory requires degradation of energy, and is made at the expense of a certain amount of negentropy, taken away from the surroundings.

The term "negative entropy" had earlier (in 1944) been used by Schrödinger in his famous book "What is life?" [98]. Specifically, he argued that "What an organism feeds upon is negative entropy". At the same time, he did not mention any other "life principle", additional to the Second Law, which would drive life and evolution.

There is no general agreement about the meaning of what negative entropy or negentropy are. Some (e.g. [99]) use them as technical terms meaning the difference of the entropy of any variable from that of a variable with normal distribution with the same mean and variance (distance to normality). However, others, in a rather metaphysical context and assuming a non-statistical definition of negentropy (e.g. [100]), see a negentropic principle governing life, biosphere, economy, etc., because these convert things which have less order, into things with more order.

Apparently, if we get rid of the disorder interpretation of entropy, we may also quit seeking a negentropic "life principle", which was never found and probably will never be. For, if we see entropy as uncertainty, we also understand that life is fully consistent with entropy maximization. Life generates new options and increases uncertainty. Compare Earth with a lifeless planet: Where is uncertainty greater? In which of the two planets a newspaper would have more events to report every day?

2.5 Final theses on entropy

The above discourse allows us to form a logical basis of a general entropic framework which can be applicable in many scientific fields including thermodynamics, geosciences and social sciences. This includes the following points.

- Entropy is a stochastic concept with a simple and general definition to be formally stated in section 2.6. Notably, according to its stochastic definition, entropy is a dimensionless quantity.
- As a stochastic concept, entropy can be interpreted as a measure of uncertainty, leaving aside the traditional but obscure and misleading "disorder" interpretation.
- The classical definition of thermodynamic entropy is not necessary; it can be abandoned and replaced by the probabilistic definition.
- Applied in thermodynamics, the thus defined entropy is the fundamental quantity, which supports the definition of all other derived ones. For example, the temperature is defined as the inverse of the partial derivative of entropy with respect to the internal energy. The entropy retains its dimensionless character in thermodynamics, thus rendering the unit of kelvin an energy unit. Notably, the extended and sophisticated study of entropy in thermodynamics can serve, after removal of the particulars pertinent to this specific field, as a paradigm for other disciplines, given that entropy is a generic concept.
- The entropy concept is equipped with the principle of maximum entropy, which states that entropy tends to take its maximum value that is allowed, given the available information about the system. The latter is incorporated in the maximization in the form of constraints. This can be regarded both as a physical (ontological) principle obeyed by natural systems, as well as a logical (epistemological) principle applicable in making inference about natural systems.
- The tendency of entropy to reach a maximum is the driving force of natural change.
- Life, biosphere and social processes are all consistent with the principle of maximum entropy as they augment uncertainty. Therefore, no additional "life principle" is necessary to explain them. Change in life and evolution are also driven by the principle of maximum entropy.

2.6 Mathematical formulation

We consider a stochastic (random) variable $\underline{x}$ (notice that we underline stochastic variables to distinguish them from regular variables) and we denote its distribution function (i.e. probability of non-exceedance) and its tail function (i.e. probability of exceedance), respectively, as:

$$F(x) := P\{\underline{x} \le x\}, \qquad \bar{F}(x) = 1 - F(x) = P\{\underline{x} > x\} \tag{1}$$

where P denotes probability. If the variable $\underline{x}$ is discrete, i.e. it can take any of the values $x_j, j = 1, \dots, \Omega$, with probability

$$P_j \equiv P(x_j) := P\{\underline{x} = x_j\} \tag{2}$$

then the sequence P_j defines its probability mass function. If the variable is continuous, i.e. it can take any real value (or a value in a subset of the real numbers), then we define the probability density function as the derivative of the distribution function:

$$f(x) := \frac{\mathrm{d}F(x)}{\mathrm{d}x} \tag{3}$$

The sequence P_j and the function $f(x)$ obey the obvious relationships:

$$\sum_{j=1}^{\Omega} P_j = 1, \qquad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1 \tag{4}$$

Any deterministic function of $\underline{x}$, $g(\underline{x})$, is a stochastic variable per se, because its argument is stochastic. The expectation of the stochastic variable $g(\underline{x})$ is defined as:

$$\mathrm{E}[g(\underline{x})] := \sum_{j=1}^{\Omega} g(x_j)\, P_j\,, \qquad \mathrm{E}[g(\underline{x})] := \int_{-\infty}^{\infty} g(x) f(x)\mathrm{d}x \tag{5}$$

for a discrete and a continuous stochastic variable, respectively. For $g(\underline{x}) = \underline{x}$, we get the mean of $\underline{x}$:

$$\mu := \mathrm{E}[\underline{x}] := \sum_{j=1}^{\Omega} x_j\, P_j\,, \qquad \mu := \mathrm{E}[\underline{x}] := \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \tag{6}$$

and for $g(\underline{x}) = (\underline{x} - \mu)^2$, we get the variance of $\underline{x}$:

$$\gamma := \mathrm{E}\left[(\underline{x} - \mu)^2\right] := \sum_{j=1}^{\Omega} (x_j - \mu)^2\, P_j\,, \qquad \gamma := \mathrm{E}\left[(\underline{x} - \mu)^2\right] := \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\mathrm{d}x \tag{7}$$

The variance is necessarily nonnegative and its square root, $\sigma := \sqrt{\gamma}$, is the standard deviation. For nonnegative variables, the ratio σ/μ, termed the coefficient of variation, is a useful dimensionless index of the variability of a system.

In the above presentation of these basic probabilistic notions, we have followed Kolmogorov's axiomatic system of probability [101,102] and we will do the same in what follows. According to this system, the definition of a stochastic variable $\underline{x}$ entails an enumeration of the basic set (the set of all possible elementary events). Hence, it reflects arbitrary choices (e.g. about units) as there are many different options for enumeration. In turn, expectations and moments depend on the option chosen. One may think of defining the function $g(\,)$ whose expectation is sought, in terms of the probability per se, i.e. $g(\underline{x}) = h(P(\underline{x}))$ for a discrete variable or $g(\underline{x}) = h\left(f(\underline{x})\right)$ for a continuous variable, where $h(\,)$ is any specified function. Among the several choices of $h(\,)$, most useful is the logarithmic function, which results in the definition of entropy.

The emergence of the logarithm in the definition of entropy follows some postulates originally set up by Shannon (1948). Assuming a discrete stochastic variable $\underline{x}$ with probability mass function $P_j \equiv P(x_j)$, which satisfies equation (4), the postulates, as reformulated by Jaynes [103, p. 347], are:

(a) It is possible to set up a numerical measure Φ of the *amount of uncertainty* which is expressed as a real number.
(b) Φ is a continuous function of P_j.
(c) If all the P_j are equal ($P_j = 1/\Omega$) then Φ should be a monotonic increasing function of Ω.
(d) If there is more than one way of working out the value of Φ, then we should get the same value for every possible way.

Quantification of postulate (d) is given, among others, in Robertson [59, p. 3] and Uffink [104, theorem 1], and is related to refinement of partitions to which the probabilities P_j refer.

From these general postulates about uncertainty, a unique (within a multiplicative factor) metric Φ results, which serves as the definition of entropy:

$$\Phi[\underline{x}] := \mathrm{E}[-\ln P(\underline{x})] = -\sum_{j=1}^{\Omega} P_j \ln P_j \tag{8}$$

We note that in classical thermodynamics, entropy is denoted by S (the original symbol used by Clausius; see section 2.1), while probability texts use the symbol H. Here Φ was preferred as a unifying symbol for information and thermodynamic entropy, under the interpretation that the two are essentially the same thing.

Extension of the above definition for the case of a continuous stochastic variable $\underline{x}$ with probability density function $f(x)$, is possible, although not contained in Shannon's (1948) original work. This extension presents some difficulties. Specifically, if we discretize the domain of x into intervals of size δx, then (8) would give an infinite value for the entropy as δx tends to zero (the quantity $-\ln P = -\ln(f(x)\,\delta x)$ will tend to infinity). However, if we involve a (so-called) *background measure* with density $\beta(x)$ and take the ratio $(f(x)\delta x)/(\beta(x)\delta x) = f(x)/\beta(x)$, then the logarithm of this ratio will generally converge. This allows the definition of entropy for continuous variables as (see e.g. [103, p. 375, 104]):

$$\Phi[\underline{x}] := \mathrm{E}\left[-\ln\frac{f(\underline{x})}{\beta(\underline{x})}\right] = -\int_{-\infty}^{\infty} \ln\frac{f(x)}{\beta(x)} f(x)\mathrm{d}x \tag{9}$$

The background measure $\beta(x)$ can be any probability density, proper (with integral equal to 1, as in equation (4)) or improper (meaning that its integral diverges). Typically, it is an (improper) Lebesgue density, i.e. a constant.

We note that most books do not include the background measure $\beta(\underline{x})$ in the definition (or set $\beta(\underline{x}) \equiv 1$) but in terms of physical consistency this is an error, because in order to take the logarithm of a quantity, this quantity must be dimensionless. While probability mass $P(x)$ in a discrete variable is indeed dimensionless, the density function has units $[f(x)] = [x^{-1}]$ and therefore we need to divide it by a quantity with same units before taking the logarithm. Even if we choose the Lebesgue measure as background, with $\beta(x) = 1/\lambda$, (constant), where λ is the unit used to measure x, still the entropy depends on the unit. It can easily be verified that if we measure x with two different units λ_1 and λ_2, the respective entropies $\Phi_1[\underline{x}]$ and $\Phi_2[\underline{x}]$ will differ by a constant:

$$\Phi_1[\underline{x}] - \Phi_2[\underline{x}] = \ln\frac{\lambda_2}{\lambda_1} \tag{10}$$

In other words, in contrast to the discrete variables where the entropy for a specified probability mass function is a unique number, in continuous variables the value of entropy depends on the assumed $\beta(x)$.

Furthermore, we note that in texts that miss the background measure in the entropy definition, the quantity that is defined in equation (9), taken with a negative sign, is named the *relative entropy* or the *Kullback–Leibler divergence*, as it measures how the density function $f(x)$ differs from $\beta(x)$.

It is easily seen that for both discrete and continuous variables the entropy $\Phi[\underline{x}]$ is a dimensionless quantity. For discrete variables it can only take nonnegative values up to a maximum value, depending on the system. For continuous variables it can be either positive or negative, depending on the assumed $\beta(x)$, ranging from $-\infty$ to a maximum value, depending on the system and, in particular, on its constraints.

If there is no constraint about the system, apart from a maximum value Ω, i.e. if the system only obeys the inequality constraint:

$$0 \leq x \leq \Omega \tag{11}$$

then, maximization of entropy results in uniformity, i.e. $P(x_j) = 1/\Omega$ or $f(x) = 1/\Omega$, while the maximum entropy is

$$\Phi[\underline{x}] = \ln \Omega\,, \qquad \Phi[\underline{x}] = \ln \frac{\Omega}{\lambda} \tag{12}$$

for the discrete and the continuous case, respectively. The former equation corresponds to the original Boltzmann's definition of entropy, a form of which has been carved on his gravestone. In the latter case, a Lebesgue background measure is assumed, i.e. $\beta(x) = 1/\lambda$.

However, a system becomes more interesting when, in addition to inequality constraints like (11), or even in absence of them, there appear equality constraints, corresponding to the information that is known about a system represented by the variable $\underline{x}$. Their formulation is typically given in the form of expectations of one or more functions $g_i(\underline{x})$ for some i:

$$\mathrm{E}[g_i(\underline{x})] = \gamma_i \Leftrightarrow \int_{-\infty}^{\infty} g_i(x) f(x) \mathrm{d}x - \gamma_i = 0 \tag{13}$$

As shown in [105], for any background measure $\beta(x)$, after incorporating the constraints to the entropy with Lagrange multipliers, the entropy maximizing density is:

$$f(x) = A\,\beta(x) \exp\left(-\sum_i b_i g_i(x)\right) \tag{14}$$

where A and b_i are parameters to be determined from the constraints of equations (4) and (13). Once $f(x)$ is determined, the maximum entropy is calculated by equation (9).

In closing the presentation of our general entropic framework, we return to the entropy definition and stress the importance of the postulate (d). This allows to separate a whole big system in partitions, requiring that, once entropy of the whole is maximized, the partial entropy in each of the partition blocks is also maximized. In turn, this enables the study of subsystems without necessarily considering the entire system. For example, we can study the economy of a country without considering all processes on Earth or on the universe.

3. Entropy maximizing distribution for constrained mean

3.1 Lebesgue background measure and the exponential distribution

In studying the material wealth in a certain society, current or past, we assume two characteristic quantities: the mean of wealth μ, which is related to the total energy available to the society [106], and an upper limit of wealth Ω, which is mainly determined by the available technology (knowhow) and thus we call it technological upper limit. We define the ratio:

$$A \coloneqq \frac{\Omega}{\mu} \geq 1 \tag{15}$$

Hence, in entropy maximization we have an equality constraint and an inequality one, i.e.:

$$\int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \mu, \qquad 0 \leq x \leq A\mu \tag{16}$$

The probability that maximizes entropy is determined from the general solution (14). Assuming a Lebesgue background measure with $\beta(x) = 1/\lambda$, with λ being a monetary unit (e.g. λ = 1 \$), after algebraic manipulations, we find the entropy maximizing probability density as:

$$f(x) = \frac{b\, \mathrm{e}^{-bx/\mu}}{\mu(1 - \mathrm{e}^{-bA})} \tag{17}$$

which is a (doubly) bounded exponential distribution (or anti-exponential if $b < 0$). The number b depends on A and is the solution of the implicit equation:

$$\frac{1}{b} - \frac{A}{\mathrm{e}^{Ab} - 1} = 1 \tag{18}$$

which renders b a function of A, $b(A)$, albeit implicitly determined. The entropy is then found to be:

$$\Phi[\underline{x}] = b(A) + \ln\left(\frac{\mu}{\lambda}\right) + \ln\left(\frac{A}{(A-1)\, b(A) + 1}\right) \eqqcolon \Phi(\mu, A) \tag{19}$$

Interesting special cases of the general solution (17) are encountered for $A = 1$ (an impulse function with all mass concentrated at $x = \mu$, representing certainty), $A = 2$ (the uniform distribution) and $A \to \infty$ (the unbounded exponential distribution). Their particular characteristics are given in Table 1 as functions of μ and in Table 2 as functions of Ω.

Accurate solutions of equation (18) can be directly calculated in terms of the auxiliary variable:

$$c \coloneqq \mathrm{e}^{Ab} \tag{20}$$

Starting with a known c, the exact solution is readily found as

$$A = \left(\frac{1}{\ln c} - \frac{1}{c-1}\right)^{-1}, \qquad b = 1 - \frac{\ln c}{c - 1} \tag{21}$$

where for $c \geq \mathrm{e}$ we get $A \geq 2, b \geq 0$, and for $c \leq \mathrm{e}$ we get $A \leq 2, b \leq 0$. Two sample solutions for $c = 2^{-3}$ and $c = 2^{3}$ are also shown in Table 1 and Table 2.

Table 1. Special (limiting) cases of the entropy maximizing distribution of equation (18) for constant mean μ and varying technological limit $\Omega = A\mu$, along with two exact solutions* (#2, #4) calculated by equation (21) for $c = 2^{-3}$ and $c = 2^{3}$, respectively.

#	$\boldsymbol{A}$	$\boldsymbol{b}$	$\boldsymbol{f(x)}$	$\boldsymbol{\Phi[\underline{x}]}$	**Distribution**
1	1	$-\infty$	$\delta(x-\mu)$	$-\infty$	Certain ($x = \mu$)
2	$\frac{21\ln 2}{24\ln 2 - 7} = 1.511$	$1 - \frac{24\ln 2}{7} = -1.377$	$\frac{0.197\mathrm{e}^{1.377x/\mu}}{\mu}$	$1 + \ln\left(\frac{\mu}{\lambda}\right) - \ln\left(\frac{2^{24/7}(24\ln 2 - 7)}{49}\right) = 0.250 + \ln\left(\frac{\mu}{\lambda}\right)$	Truncated anti-exponential
3	2	0	$1/2\mu$	$\ln 2 + \ln\left(\frac{\mu}{\lambda}\right)$	Uniform
4	$\frac{21\ln 2}{7 - 3\ln 2} = 2.958$	$1 - \frac{3\ln 2}{7} = 0.703$	$\frac{0.803\mathrm{e}^{-\frac{0.703x}{\mu}}}{\mu}$	$1 + \ln\left(\frac{\mu}{\lambda}\right) - \ln\left(\frac{2^{24/7}(7 - 3\ln 2)}{49}\right) = 0.922 + \ln\left(\frac{\mu}{\lambda}\right)$	Truncated exponential
5	$\to\infty$	1	$\mathrm{e}^{-x/\mu}/\mu$	$1 + \ln\left(\frac{\mu}{\lambda}\right)$	Unbounded exponential

* Numerals are rounded to three decimal digits.

Table 2. Special (limiting) cases of the entropy maximizing distribution of equation (18) for constant technological limit Ω and varying mean $\mu = \Omega/A$, along with two exact solutions* (#2, #4) calculated by equation (21) for $c = 2^{-3}$ and $c = 2^{3}$, respectively.

#	$\boldsymbol{A}$	$\boldsymbol{b}$	$\boldsymbol{f(x)}$	$\boldsymbol{\Phi[\underline{x}]}$	**Distribution**
1	1	$-\infty$	$\delta(x-\Omega)$	$-\infty$	Certain ($x = \Omega$)
2	$\frac{21\ln 2}{24\ln 2 - 7} = 1.511$	$1 - \frac{24\ln 2}{7} = -1.377$	$\frac{0.297\mathrm{e}^{2.079x/\Omega}}{\Omega}$	$1 + \ln\left(\frac{\Omega}{\lambda}\right) - \ln\left(\frac{3(2^{24/7})\ln 2}{7}\right) = -0.163 + \ln\left(\frac{\Omega}{\lambda}\right)$	Truncated anti-exponential
3	2	0	$1/\Omega$	$\ln\left(\frac{\Omega}{\lambda}\right)$	Uniform
4	$\frac{21\ln 2}{7 - 3\ln 2} = 2.958$	$1 - \frac{3\ln 2}{7} = 0.703$	$\frac{2.377\mathrm{e}^{-2.079x/\Omega}}{\Omega}$	$1 + \ln\left(\frac{\Omega}{\lambda}\right) - \ln\left(\frac{3(2^{24/7})\ln 2}{7}\right) = -0.163 + \ln\left(\frac{\Omega}{\lambda}\right)$	Truncated exponential
5	$\to\infty$	1	$\delta(x)$	$-\infty$	Certain ($x = 0$)

* Numerals are rounded to three decimal digits.

In the typical case where c is unknown we need to solve equation (18) numerically, which is not too difficult. Alternatively, we can use the following very good approximation for $A \geq 2$:

$$b = 1 - (1 - 1.37\mathrm{e}^{-0.89})\,\mathrm{e}^{2(2-A)} - 1.37\,\mathrm{e}^{0.89(1-A)} \tag{22}$$

Furthermore, a very good approximation for the entropy $\Phi[\underline{x}]$ which does not contain b at all, is:

$$\Phi[\underline{x}] = 1 + \ln\left(\frac{\mu}{\lambda}\right) - (1 - \ln 2 - 0.373\mathrm{e}^{-1})\,\mathrm{e}^{2(2-A)} - 0.373\,\mathrm{e}^{(1-A)} \tag{23}$$

Approximations (22) and (23) were found by an extensive numerical investigation and are optimized for $A \geq 2$. For $A < 2$ we can exploit the following symmetry relationships:

$$b(A) = -\frac{1}{A-1} b\left(\frac{A}{A-1}\right), \qquad \Phi(\mu, A) = \Phi\left((A-1)\mu, \frac{A}{A-1}\right) \tag{24}$$

where it is readily seen that if $A < 2$ then $A/(A-1) > 2$.

In addition to their tabulated form (Table 1), the characteristic density functions are also depicted in graphical form in Figure 5. Likewise, the density functions of Table 2 are depicted in Figure 6. Furthermore, Figure 7(a) shows the achieved maximum entropy for a constant mean $\mu = 1$, as a function of the technological limit Ω. For small and moderate values of the technological limit, the entropy is an increasing function of Ω but beyond $\Omega \approx 5$, it reaches a virtually constant value. Likewise, Figure 7(b) shows the maximum entropy for a constant technological limit $\Omega = 2.958$ (this value corresponds to case #4 of Table 2) as a function of the mean μ. Initially, for small μ, the entropy increases; it takes a maximum value for $\mu = \Omega/2$ and then decreases following a symmetric pattern. However, in the case that technology offers unlimited opportunities (infinite technological limit), also depicted in Figure 7(b), the increase of entropy with μ is continuous. We can thus say in conclusion that for $\Omega > 2\mu$ the entropy increases both with the mean μ and the technological limit Ω. In this respect the entropy constitutes a measure of society's wealth (see also [106]).

One could say that the mean μ is more representative, as a measure of wealth, than the entropy Φ. We do not have substantial objection on this. However, we prefer to use entropy for two reasons: (a) because it is connected with the logarithm of the mean, and intuitively this is a better quantification of the wealth, and (b) because it quantifies the options, and as seen in the first motto in the beginning of the paper, the quantification of options is more pertinent as a measure of average wealth.

Apparently, life offers much more options than material wealth and one could choose to pursue different opportunities (e.g. intellectual), snubbing material wealth, as formulated by Bias of Priene in the second motto in the beginning of the paper. Certainly, the focus of the paper is on material wealth but we should keep in mind that seeking material wealth is just one of the options. (For example, we would not like to interchange our lives with any one of the persons whose income is depicted in Figure 9 to be discussed below.)

While for constant background density equal to the inverse of the monetary unit (i.e. $1/\lambda$ with λ equal e.g. to 1 \$) the entropy provides a measure of society's wealth, if we change the background measure to the value $1/\mu$, where μ is the mean income, the thus calculated entropy is a measure of inequality. Calling the latter quantity *standardized entropy* and denoting it as $\Phi_\mu[\underline{x}]$, from equation (10) we get

$$\Phi_\mu[\underline{x}] = \Phi[\underline{x}] - \ln\frac{\mu}{\lambda} \tag{25}$$

This has been recently introduced as an index of inequality by Sargentis et al. [106] (albeit denoted as $\Delta\Phi[\underline{x}]$). The quantity $\Phi_\mu[\underline{x}]$ cannot exceed a maximum value of 1, corresponding to an exponential distribution. A value smaller than 1 usually indicates smaller inequality with respect to the exponential distribution. But, as we will see in section 3.2, there are cases where it indicates higher inequality and hence the value of $\Phi_\mu[\underline{x}]$ should be accompanied with a second inequality index in order to decide whether the inequality is lower or higher. As we will see below a simple and appropriate additional index is the coefficient of variation σ/μ.

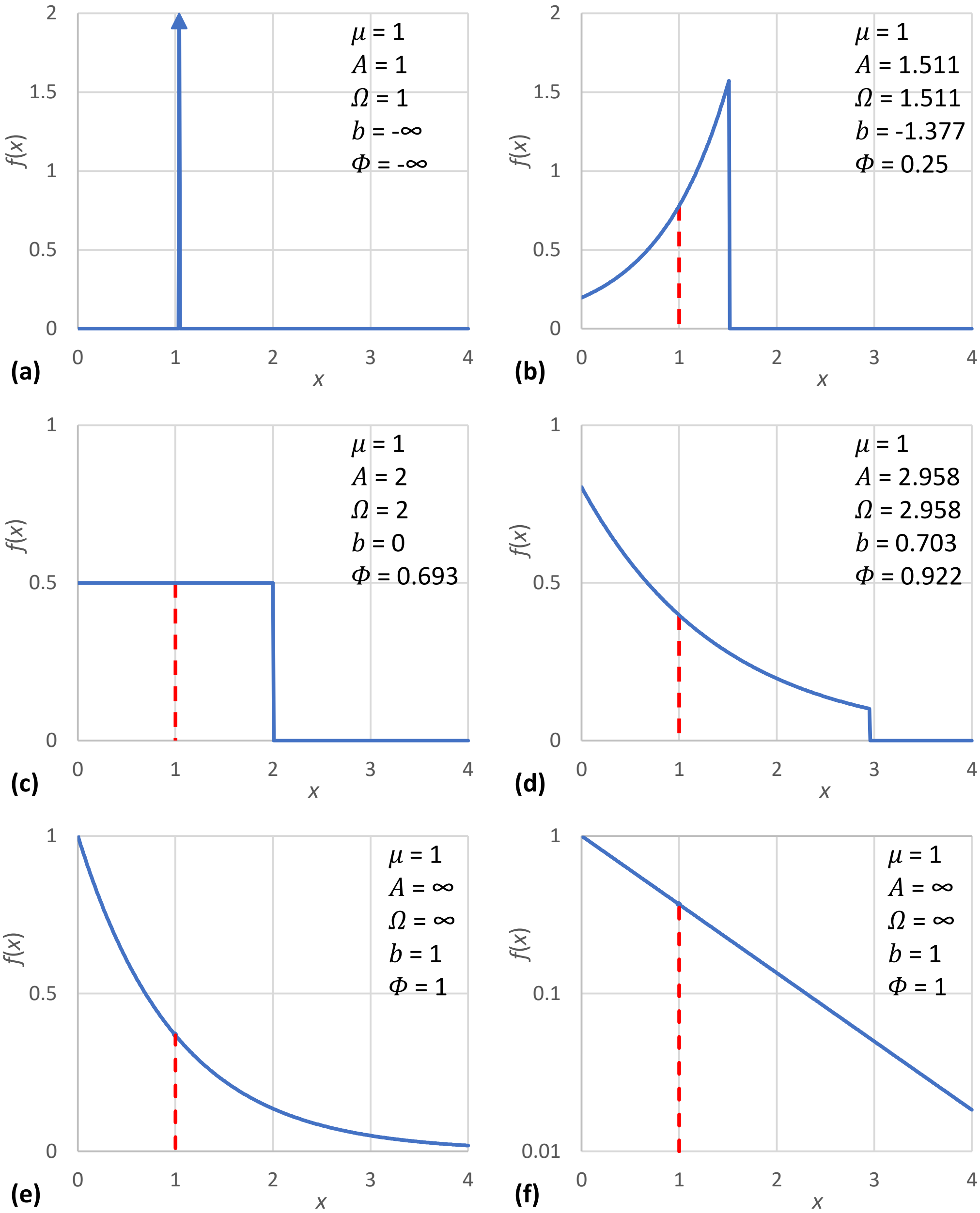


Figure 5. Entropy maximizing probability density functions of Table 1 for mean $\mu = 1$. Panels (**a**) to (**e**) correspond to cases 1 to 5 of Table 1 while panel (**f**) is same with (e) except that the vertical axis is logarithmic. In each panel the mean is depicted as a red dashed line.

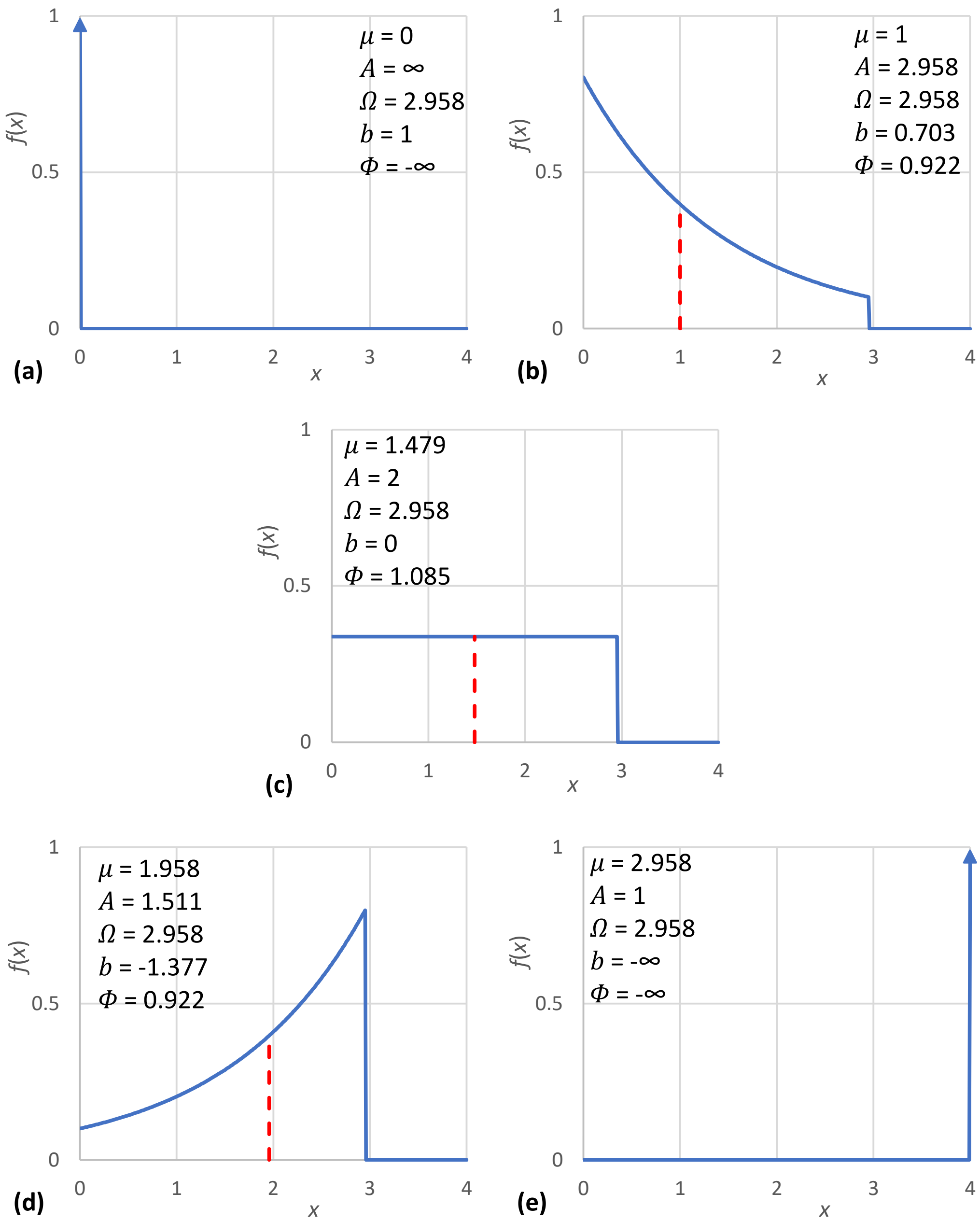


Figure 6. Entropy maximizing probability density functions of Table 2 for mean technological limit $\Omega = 2.958$ (corresponding to $\mu = 1$ for case #4 of Table 2). Panels (**a**) to (**e**) correspond to cases 5 to 1 of Table 2 (in reverse order, so that the mean is increasing from 0 in panel (a) to Ω in panel (e)). In each panel the mean is depicted as a red dashed line.

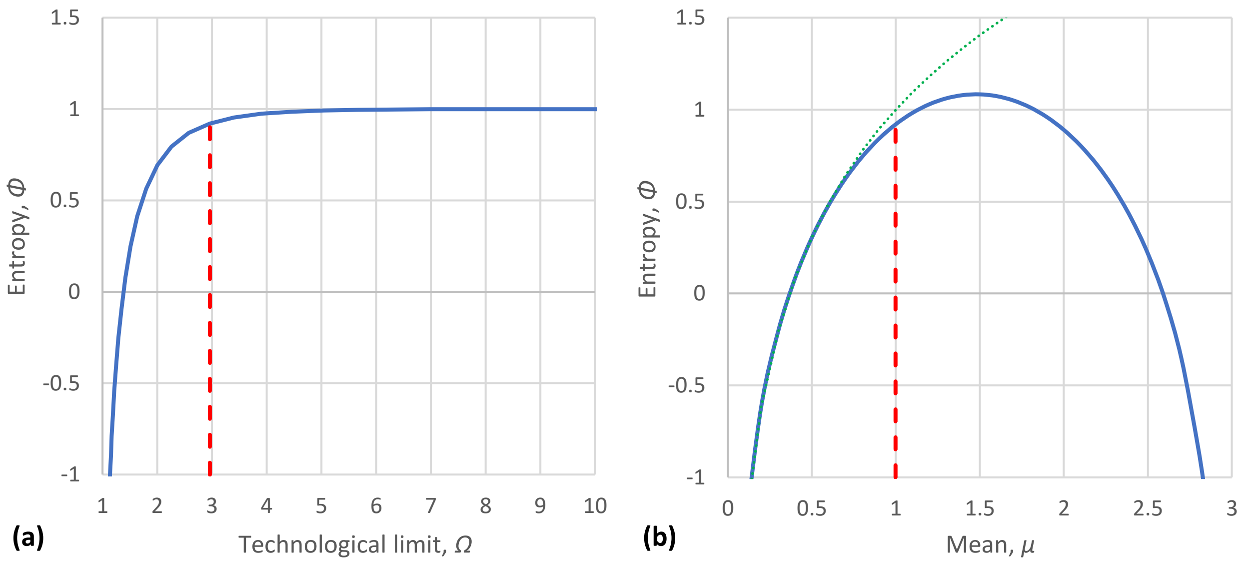


Figure 7. Maximum entropy as a function of **(a)** the technological limit Ω for constant mean $\mu = 1$ and **(b)** of the mean μ for constant technological limit $\Omega = 2.958$. The red dashed lines correspond to the identical cases in the two panels. The green dotted line in panel (b) depicts the maximum entropy for infinite technological limit.

3.2 Hyperbolic background measure and the Pareto distribution

Coming again to the quantification of material wealth, we recall that the obtained density function and maximum entropy, would be different if we chose a different background measure. In particular, if we choose a hyperbolic background measure [68,105], i.e.

$$\beta_{\mathrm{H}}(x) = \frac{1}{\lambda + x} \tag{26}$$

leave x unbounded, and constrain a generalized mean, consistent with the chosen background measure, then the entropy maximizing distribution becomes Pareto, with density

$$f(x) = \frac{1}{\lambda_1}\left(1 + \xi\frac{x}{\lambda_1}\right)^{-\frac{1}{\xi}-1}, \qquad x \geq 0 \tag{27}$$

where $\xi \coloneqq 1/b_1$ (the parameter in equation (14), known as the tail index) and $\lambda_1 \coloneqq \lambda\xi$. It can be seen (by taking the limit) that, as $\xi \to 0$, the Pareto density (equation (27)) tends to the unbounded exponential one (case #5 in Table 1) with $\mu = \lambda_1$.

We note that the density in equation (27) is often called the Pareto Type II or Lomax distribution, while the name Pareto distribution (more precisely, Pareto Type I distribution) is used for the case where the 1 in the parenthesis of equation (27) is neglected; in that case x cannot take values smaller than λ_1/ξ. The differences in the two cases are negligible for large x.

The distribution, else known as the "Pareto's law", is named after Vilfredo Pareto who first proposed it in 1896 while analysing data on the distribution of wealth and fitting a straight line on the logarithm of the number of people N_x, whose net worth or income exceeds x, and the logarithm of x [107]. A detailed historical account of his discovery is given by Mornati [108], where it can be seen that Pareto developed both types I and II of the distribution. In fact, it was not the power-law behaviour of the distribution that impressed Pareto. As evident from his celebrated book *Manual of Political Economy*, first published in 1906, he was rather puzzled by its asymmetric shape, which is also shared by the exponential distribution. Specifically, he drew a qualitative shape of the distribution in Figure 54 of his book and he remarked [109, p. 195]:

> *The shape of the curve* […] *of Figure 54, which is derived from statistics, does not correspond by any means to the curve of errors* [(the normal distribution)], *i.e., to the shape the curve*

would have if the acquisition and preservation of wealth depended only on chance. Moreover, statistics reveal that the curve […] varies very little in time and space; different nations at different times have very similar curves. There is thus a remarkable stability in the shape of this curve. […] There is a certain minimum income […] below which men cannot descend without dying of poverty and hunger.

It is remarkable that Pareto regarded that "chance" in only connected with the normal distribution—an idea that is not consistent with reality. If we accept that what Pareto regarded as "chance" can be represented by the principle of maximum entropy, then it is true that there is consistency of the latter principle with the normal distribution. However, this occurs when the second moment of $\underline{x}$ is constrained to a constant value (as e.g. in the kinetic energy of a number of molecules) and certainly such a constraint is not applicable to income. Instead, as we have seen, reasonable constraints for the income necessarily result in asymmetric distributions, exponential or Pareto. In this respect, the Pareto distribution is applied and has become popular in financial analysis, in its original form of after generalization [110].

Coming back to the mathematical details, the mean and coefficient of variation of the Pareto distribution are

$$\mu = \frac{\lambda_1}{1-\xi} = \frac{\lambda\xi}{1-\xi}, \qquad \frac{\sigma}{\mu} = \frac{1}{\sqrt{1-2\xi}} \tag{28}$$

Assuming that both should have finite values, a restriction is imposed for ξ, i.e. $0 \le \xi < 1/2$. The maximized entropy for the hyperbolic background measure is

$$\Phi_{\mathrm{H}} = 1 + \ln\xi = 1 - \ln(1 + \lambda/\mu) \tag{29}$$

while if, for the sake of compatibility, we also calculate the entropy for the Lebesgue measure $\beta(x) = 1/\lambda$, the entropy is

$$\Phi = 1 + \xi + \ln\xi = 1 + \frac{1}{1+\lambda/\mu} - \ln(1 + \lambda/\mu) \tag{30}$$

The proofs are omitted for both cases. It can also be shown (again the proof is omitted) that the following inequality holds true:

$$\Phi_{\mathrm{H}} \le \Phi \le 1 + \ln\left(\frac{\mu}{\lambda}\right) \tag{31}$$

where the rightmost term is the entropy of the unbounded exponential distribution. The three quantities become equal as $\mu/\lambda \to 0$ (or, equivalently, as $\xi \to 0$).

It is relevant here to discuss the so-called "Pareto principle" or the "80/20 rule" referred to in economic papers (e.g. [111]) and suggesting that about 80 percent of wealth is concentrated in about 20 percent of a population. The mathematics for this principle is the following. Let $\mu_{x\ge c} \coloneqq \int_c^\infty x f(x) \mathrm{d}x$ be the mean of $\underline{x}$ conditional on $\underline{x} > c$. The "principle" states that for $\mu_{x\ge c}/\mu = 0.8$, $\overline{F}(c) = 0.2$.

It can be shown that for the Pareto distribution the following relationship holds true:

$$\frac{\mu_{x\ge c}}{\mu} = \frac{\left(1 + \left(\overline{F}(c)^{-\xi} - 1\right)/\xi\right)\overline{F}(c)}{1-\xi} \tag{32}$$

and it can be readily verified that the condition $\mu_{x\ge c}/\mu = 0.8, \overline{F}(c) = 0.2$ is met when $\xi = 0.253$. For $\xi \to 0$ the Pareto distribution reduces to the exponential distribution and equation (32) reduces to

$$\frac{\mu_{x\ge c}}{\mu} = \left(1 - \ln\overline{F}(c)\right)\overline{F}(c) \tag{33}$$

From the latter, it is easy to verify that the condition $\mu_{x\ge c}/\mu = 0.8$ is satisfied when $\overline{F}(c) = 0.439$, substantially higher than 0.2.

When the articles refer to the "Pareto principle", they usually cite Pareto's Manual [109]. However, our own perusal did not show that the "principle" is referred to or implied by his book. On the contrary, it appears that this name was accidentally given in 1951 by Juran [112], as explained by himself in his paper entitled "The non-Pareto principle; mea culpa", where he claims that it was himself who introduced the principle, also using the name "vital few and trivial many" [113]. For these reasons, in the following we will call this "principle" the "80/20 rule" and we will see that real world data on income do not satisfy it.

A graphic comparison of the Pareto density with the exponential one is given in Figure 8. In addition, the two-parameter gamma density, whose behaviour is opposite to Pareto, is also plotted in the figure. The gamma density, with a chosen shape parameter 2, is $f(x) = xe^{-x/a}/a^2$ and has mean $\mu = 2a$, coefficient of variation 0.5 and entropy

$$\Phi = 1 + \gamma + \ln a/\lambda = 1 + \gamma - \ln 2 + \ln(\mu/\lambda) = 0.884 + \ln(\mu/\lambda) \quad (34)$$

where $\gamma = 0.5772\ldots$ is the Euler's constant. For the comparison we use the Lebesgue measure $\beta(x) = 1/\lambda$. In all three distributions, the domain of x is $[0, \infty)$ and the mean is $\mu = \lambda = 1$. The entropy maximizing distribution is exponential, which has entropy $\Phi = 1$. The gamma distribution has scale parameter $a = 0.5$ and entropy $\Phi = 0.884 < 1$. The tail index of the Pareto distribution was chosen so that it have the same entropy with the gamma distribution, $\Phi = 0.884$ and mean $\mu = 1$; these are obtained when $\xi = 0.408$ and $\lambda_1 = 0.592$.

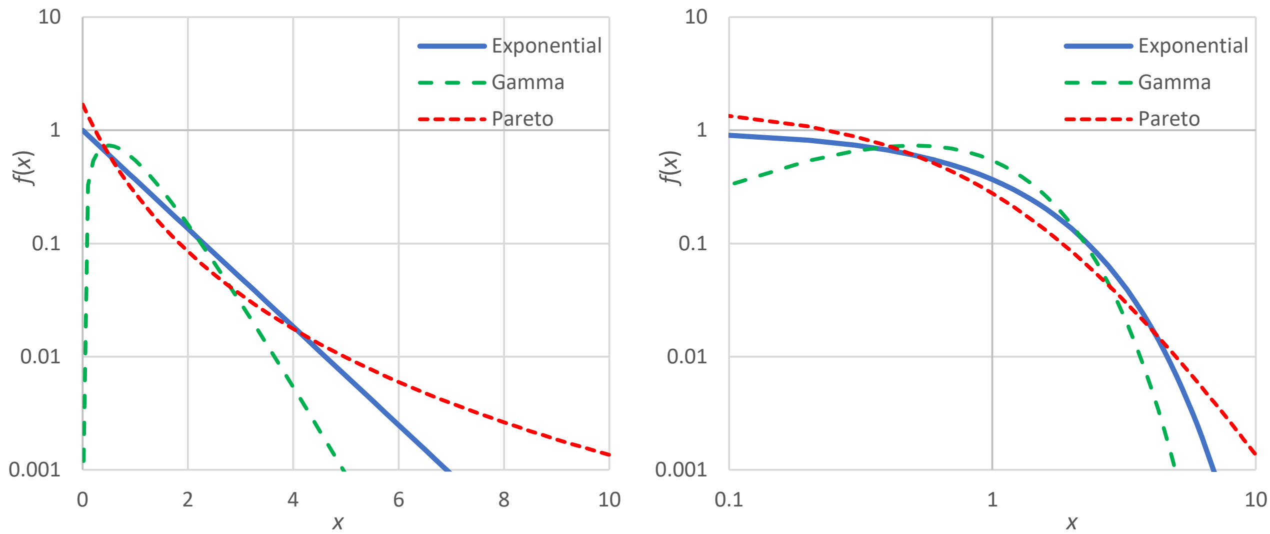


Figure 8. Comparison of the entropy maximizing distribution (for Lebesgue measure $\beta(x) = 1/\lambda$) with two other distributions, in terms of their probability density functions $f(x)$.The domain of x is $[0, \infty)$ and in all three distributions the mean is $\mu = \lambda = 1$. The entropy maximizing distribution is exponential with entropy $\Phi = 1$, and the other two are gamma with shape parameter 2 and Pareto with tail index $\xi = 0.408$. Both the gamma and the Pareto distributions have the same entropy, $\Phi = 0.884 < 1$, but they have different coefficients of variation, 0.5 and 2.325, respectively. In both panels, $f(x)$ is plotted in logarithmic axis, while the variable x is plotted in linear axis in the left panel and logarithmic axis in the right panel. Notice in the left panel that the exponential density has constant slope -1, while, the gamma density has slope -2 for large x. Likewise, in the right panel the Pareto density has constant slope $-1 - 1/\xi = -3.453$.

While the Gamma and the Pareto distributions have the same mean and the same entropy, they have different behaviour. The former favours (i.e., increases the frequency of) the moderate values of x (say, populates more the "middle class"), while the latter favours the extremes (say, diminishes the presence of the "middle class", and populates more the "poor" and the "very rich"). These differences cannot be captured by entropy alone as the entropy in the two cases is the same. One could think of using the concept of divergence from the exponential distribution, which is easy as it only needs to set the background density equal to that of the exponential distribution, i.e. $\beta(x) = \exp(-x)$ (and change the sign in the right-hand side of equation (9)). But again, this does not help, as it yields almost the same divergence: 0.116 and 0.111 for the gamma and the Pareto case, respectively.

Thus, to quantify the differences in the two cases we have to use a different metric and the simplest is the coefficient of variation, whose values are 0.5 and 2.325, for the gamma and the Pareto case, respectively.

Notice in Figure 8 that in the left panel the vertical axis is logarithmic, while in the right panel both axes are logarithmic. The left is better to visualize an exponential tail of a distribution, which appears as a straight line for large x. This is the case for both the exponential and the gamma distribution, with their slopes being -1 for and -2, respectively. The right panel better visualizes the power-law tail of a distribution, as in the Pareto case, which appears also as a straight line for large x with constant slope $-1-1/\xi = -3.453$.

3.3 Empirical investigation

In natural (e.g. geophysical) processes, both exponential and power-type tails appear, which means that the appropriate background measure, Lebesgue or hyperbolic, may differ in entropy maximization for different processes [105]. In socioeconomic processes there is no accumulated evidence about which the appropriate measure is. The most telling evidence about the type of the distribution and, hence, the appropriate background measure, is obtained by studying the distribution tail.

To study the tail, we do not need to examine the entire population, i.e. the entire range of the variable x. It suffices to examine the behaviour above a certain threshold x_0 and in particular the conditional tail function:

$$\bar{F}(x|\underline{x} > x_0) = P\{\underline{x} > x|\underline{x} > x_0\} = \frac{\bar{F}(x)}{\bar{F}(x_0)}, \qquad x \geq x_0 \tag{35}$$

An important property of both the exponential and the Pareto distribution (not shared with other common distributions) is that, if the variable is shifted by x_0, i.e. $\underline{y} \coloneqq \underline{x} - x_0$, then the distribution is preserved, with only the scale parameter changed. This implies that the coefficient of variation of $\underline{x} - x_0$ for the values of $\underline{x}$ exceeding x_0 is the same as that of $\underline{x}$ for the entire population.

In order to empirically study the tail of income distribution, we used data for the net worth of richest people of the world (billionaires) and the evolution thereof. We located the database which is referenced in the Forbes list [114] for the years 1996 to 2018. We evaluated these data with the Wayback Machine [115] and we found that amendments were needed for the years 1997, 2014, 2015, which we did.

For the years 2019, 2020 and 2021, we retrieved data from Bloomberg using again the Wayback Machine [116]. Specifically, the dates we located with the Wayback Machine were 8 March 2019, 2 January 2020 (just before outbreak of the Covid pandemic) and 8 June 2021; these days allow us to study how the Covid pandemic influenced the wealth of the richest people.

Each year's list contains a varying number of billionaires, with an average number of 860. The dataset of all years contains about 5000 names of billionaires. Many of them appear in the list for several years. By subtracting the total net worth of a person for each year from that of the previous year, we found an approximation of the person's annual net income. Then we sorted the list of net income in each year in decreasing order and we took the highest 100 of them in each year, setting the threshold x_0 equal to annual net income of the 100th person in each year's list. We further processed those years that contained at least 100 persons with positive annual net income—a total of 19 years.

Denoting $x_{(i:n)}$ the net income of the person ranked i in a list of n ($i = 1, \dots, n$, where in our case $n = 100$) the unbiased estimate of the probability of exceedance of the value $x_{(i:n)}$ is [105]:

$$\hat{\bar{F}}(x_{(i:n)}|\underline{x} > x_0) = \frac{i}{n+1} \tag{36}$$

The sequences of the $\hat{\bar{F}}(x_{(i:n)}|\underline{x} > x_0)$ for each year are depicted in Figure 9. In the left panel, in which the horizontal axis is linear and the vertical logarithmic, the exponential tails appear as straight lines. In contrast, in the right panel, in which both axes are logarithmic, the power-law (Pareto) tails appear as straight lines.

Inspection in both panels shows that both exponential and power-law tails appear, with the former being more common than the latter. Characteristically, in the year with the lowest earnings, 2002, the tail is clearly exponential, while in the year with the highest earnings, 2021, the tail is clearly of power-law type. It is relevant to note that year 2021, which was the most anomalous as the Covid 19 pandemic negatively affected the economy globally, also brought the historically highest profits to the richest. This is also seen in Figure 10, where the average of the 100 highest incomes in 2021 is unprecedented and several times higher than the average of the previous years.

Figure 10 also depicts the evolution of the coefficient of variation of the high incomes through the years with available data. We observe that the coefficient is consistently higher than unity, the value corresponding to the exponential distribution, but in most years the deviation from 1 is small. This justifies, at least as an approximation, the assumption of the exponential distribution as the norm, with the Pareto distribution being more common in anomalous periods. Put it otherwise, the evidence from data does not exclude the hypothesis of the exponential distribution and, hence, of entropy maximization with a Lebesgue background measure. Therefore, we will use this hypothesis as the main explanatory tool for what the normal (and hence sustainable) tendency in the distribution of wealth is. We will provide additional evidence on the plausibility of the hypothesis in section 4, where we will also discuss the forces that lead to deviation from the normal tendency.

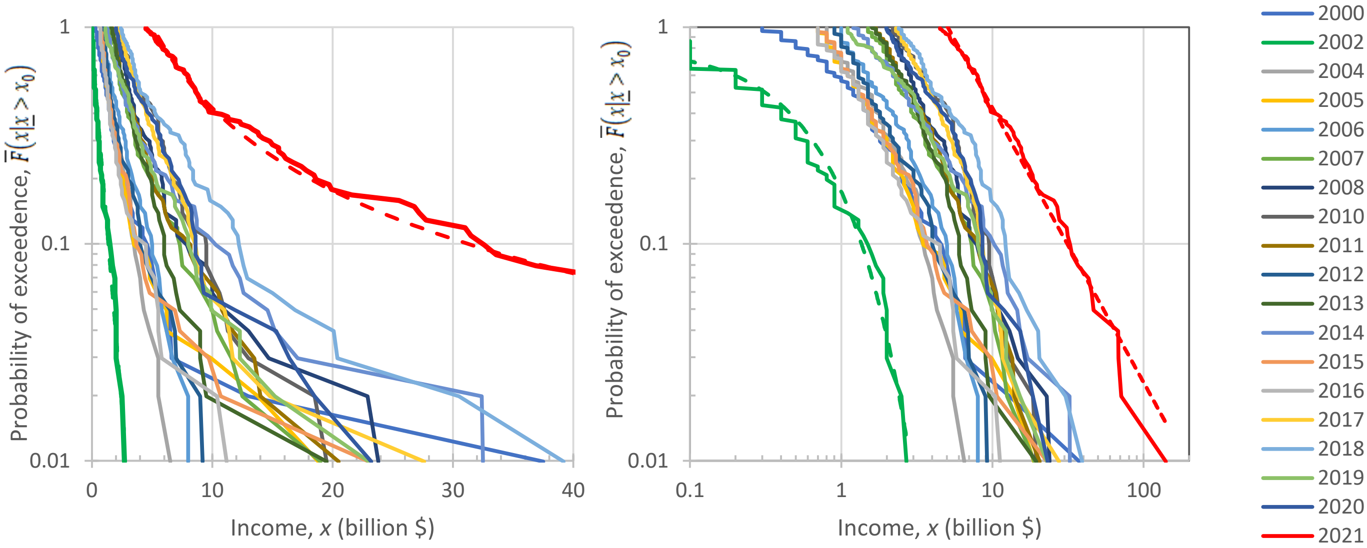


Figure 9. Conditional probability of exceedance of the annual income of the richest persons in the world for the indicated years. The income per person was found by subtracting the total net worth of a year from that of the previous year. For the years 2002 (lowest average income) and 2021 (highest average income) exponential and power-law trends, respectively, are also plotted with dashed lines of the same colour (where in the left panel the green dashed line for 2002 is indistinguishable from the continuous line). In both panels, the probability of exceedance is plotted in logarithmic axis, while the income x is plotted in linear axis in the left panel and logarithmic axis in the right panel.

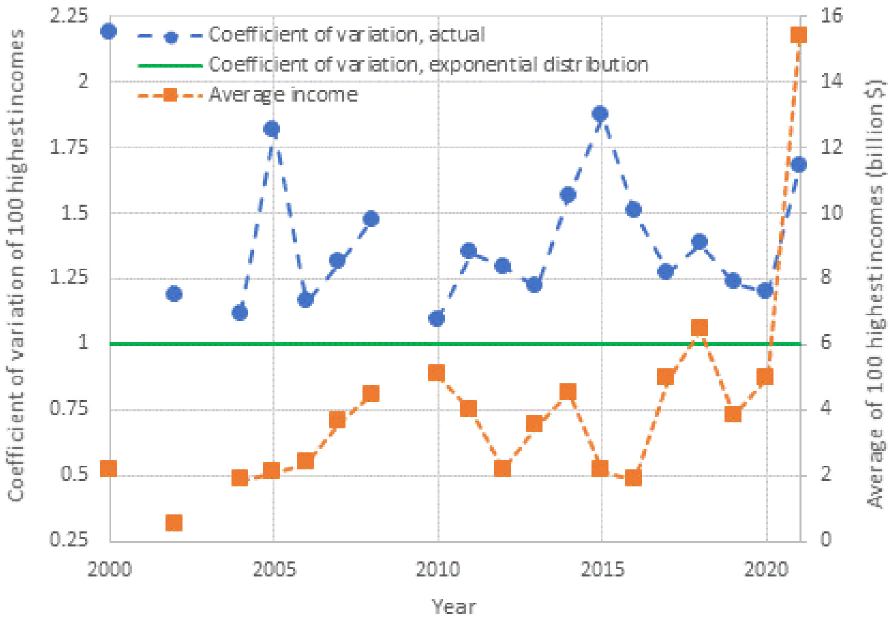


Figure 10. Average and coefficient of variation of the 100 highest incomes in the world per year. The coefficient of variation is calculated for the difference $\underline{x} - x_0$, where x_0 is the 100th highest income value, so that it be representative for the entire population (see explanation in text).

4. Application to societies' income distribution

4.1 From the ancient classless society to modern stratified societies

As suggested in the recent paper by Sargentis et al. [106], in prehistoric societies the wealth could be measured in terms of the available energy per person. Following this idea, we identify this with the income x and we set $x = 0$ to represent the energy that assures covering of the basic bodily energy needs (through food intake), keeping a human alive. Here it is relevant to note that, while energy in a closed natural system is conserved, the energy available to human individuals or societies varies and it has substantially been increased with the development of civilization.

With absence of technology (i.e. with technological limit $\Omega = 0$), the value $x = 0$ is the only option and the entropy is $-\infty$. The single allowed option and the $-\infty$ entropy, as depicted in Figure 11(a), signifies a classless society. Most probably, this corresponds to the so-called hunter-gatherer society, admired by the Marxist literature as the ancient communal ownership [117, p. 44] and aspired also for the future in a modern form.

The notion remains popular even today (Figure 12) and, strikingly, it has been regarded as a basis for real personal freedom [117,118], despite corresponding to entropy of minus infinity. Apparently, Marx and Engels were faithful to the deterministic scientific paradigm of their era, which they attempted to transplant to history and sociology. They could not have been aware of the modern concept of entropy. The popularity of their ideas even today reflects the fact that the deterministic paradigm remains quite strong. In it, entropy has no place, let alone its connection with freedom.

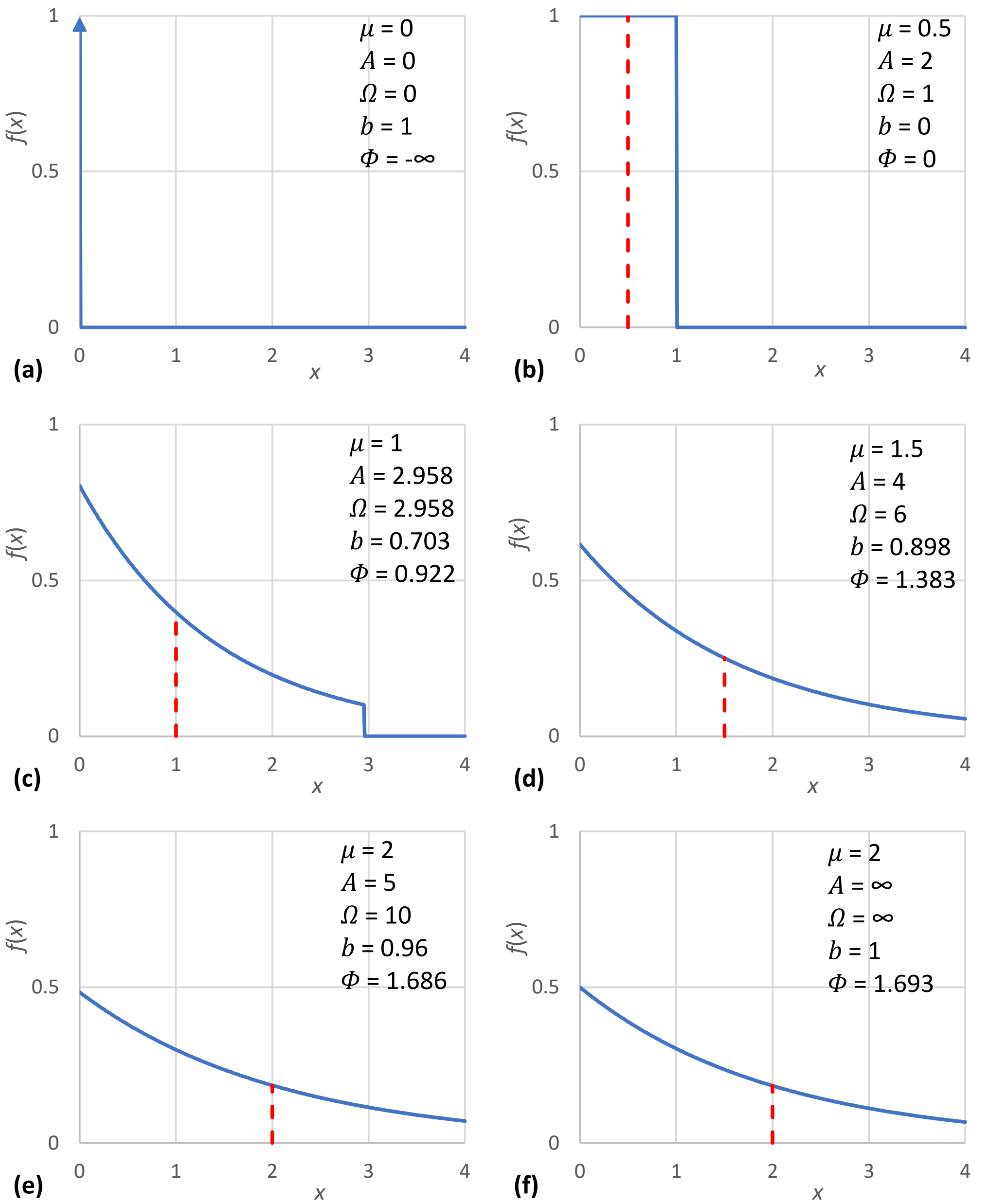


Figure 11. A possible development of human wealth from prehistory to modern times: From the primitive classless society (**a**) to uniform distribution of wealth (**b**) and to increasing diversity of wealth, inequality and stratification (**c**)-(**e**). Panel (**f**), in which the technological limit is infinite, is similar to panel (e), illustrating the fact that if the technological limit is large enough, its effect can be neglected.

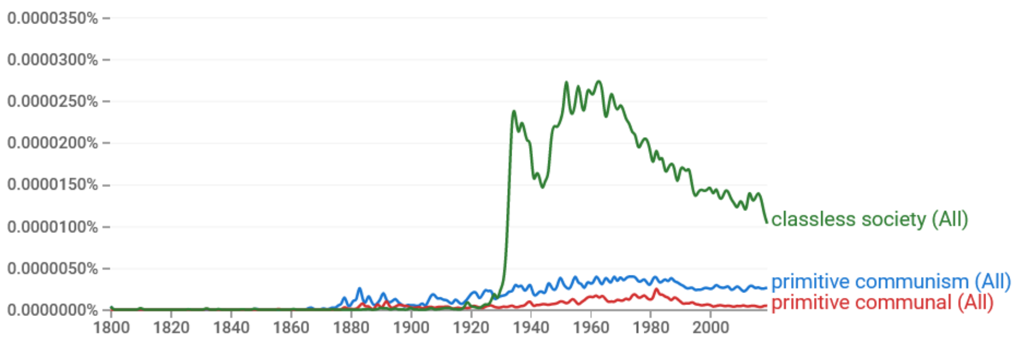


Figure 12. Frequency of appearances of the indicated phrases in Google Books [83,84].

Undoubtedly, the primitive societies developed some technologies in terms of stone tools, and therefore the technological limit $\Omega = 0$ depicted in Figure 11(a) is merely a simplification. Nonetheless, the technological revolutions that really advanced the technological limit to a higher order of magnitude were the domestication of animals (allowing use of energy additional to that of human muscles) and the invention of agriculture. At later stages, as knowledge was developed, the technological limit Ω was increased. The mean wealth μ also increased. It is plausible to assume that the rate of increase of μ always followed that of Ω. Based on this assumption, the remaining panels of Figure 11 have been constructed, to represent various phases of human wealth development.

For example, at some phase of the development, the mean wealth μ was half the technological limit Ω (Figure 11(b)). At this phase, entropy maximization suggests that the wealth was uniformly distributed among people. Uniformity means that poor and rich were equally probable—not that the wealth is equally distributed among people. When the technological limit increased more than twice the mean wealth, the distribution became (bounded) exponential. This means that the poor are more and the rich fewer—but richer than before (Figure 11(c-e)). But careful inspection of the graphs shows that it is not only the richest who become richer as technology evolved. The poor also became fewer and the curves $f(x)$ moved to the right, i.e., everybody, on the average, became richer thanks to technological evolution. At some point of evolution, when the technological limit became very high (Figure 11(e)) its effect on the wealth distribution became negligible. This also applies to modern societies, in which we can totally neglect this effect, replacing Figure 11(e) ($\mu = 2, \Omega = 10$) with Figure 11(f) ($\mu = 2, \Omega = \infty$). This is consistent with what we have already discussed in Figure 7(a).

4.2 The elites' role

It is reasonable to assume that the economic elites pursue a greater share of the community wealth. In this respect, their function can be twofold. On the one hand, they advance both the technological limit and the average wealth. On the other hand, they tend to modify the distribution of income from exponential to Pareto, thus increasing the frequency of the poor and diminishing the middle class for their own advantage (Figure 8). As has been already discussed in section 3.3, a persuasive illustration of this is the super profit of economic elites during the recent anomalous period of pandemic.

The means to increase elites' profits certainly include political power and, more recently, an attitude to control the world [e.g. 89]. Their endeavour becomes more efficient and acceptable by the society by several means they use, such as by overstating existing or non-existing threats, and then by presenting themselves as philanthropists (e.g. by funding nongovernmental organizations dealing with these threats) and world saviours [82] (see also [119]). Apparently, if they succeed in controlling the world, this will decrease entropy and hence delimit freedom. In turn, this will lead to decadence, whose signs are already visible in the Western World (cf. [120]).

4.3 Income redistribution in organized societies

One of the important roles of a state reflecting an organized society is the redistribution of income and wealth through their transfer from some individuals to others by means of several mechanisms such as taxation, public services, land reform, monetary policies, and other. Such means contrast those of the elites and aim to decrease poverty and social inequality. Here we examine one of the mechanisms, i.e. taxation, by means of a simplified toy-model example, which illustrates how redistribution affects the entropy maximizing exponential distribution.

In our toy model, we assume that the original income $\underline{x}$ follows the exponential distribution with parameter λ (equal to the mean μ) and that the tax rate p increases with the original income, according to the function:

$$p(x) = \begin{cases} 0 & x \leq x_0 \\ p_u & x \geq x_u \\ p_u \dfrac{x - x_0}{x_u - x_0}, & \text{otherwise} \end{cases} \quad (37)$$

where x_0 is a low value of income, denoting the starting point of taxation, and x_u is a high value of income, beyond which the tax rate takes a constant value p_u. The tax amount will be $\underline{y} = p(\underline{x})\underline{x}$, the income minus tax $\underline{w} = \underline{x} - \underline{y} = \left(1 - p(\underline{x})\right)\underline{x}$ and the redistributed income will be $\underline{z} = \underline{w} + a\mu_y = \left(1 - p(\underline{x})\right)\underline{x} + a\mu_y$, where $\mu_y \coloneqq E[\underline{y}]$ is the average tax and $a < 1$ is the tax fraction that is returned to all people at equal shares. In order for the relationship of $\underline{w}$ and $\underline{x}$ to be monotonically increasing the following inequalities should hold:

$$p_u < 0.5, \qquad x_u > \frac{1 - p_u}{1 - 2p_u} x_0 \quad (38)$$

Using probabilistic algebra on these variables we find that the probability density of $\underline{w}$ is:

$$f_w(w) = \begin{cases} \dfrac{1}{\lambda} \exp\left(-\dfrac{w}{\lambda}\right), & 0 \leq w \leq x_0 \\ \dfrac{x_u - x_0}{\lambda\sqrt{D(w)}} \exp\left(-\dfrac{x_u - (1 - p_u)x_0 - \sqrt{D(w)}}{2\lambda p_u}\right), & x_0 \leq w \leq w_u \\ \dfrac{1}{\lambda(1 - p_u)} \exp\left(-\dfrac{w}{\lambda(1 - p_u)}\right), & w \geq w_u \end{cases} \quad (39)$$

where

$$w_u \coloneqq (1 - p_u)x_u, \qquad D(w) \coloneqq (x_u - (1 - p_u)x_0)^2 - 4(x_u - x_0)p_u w \quad (40)$$

which allows finding the mean values:

$$\mu_w = \lambda - \frac{\lambda p_u}{x_u - x_0}\left(e^{-\frac{x_0}{\lambda}}(x_0 + 2\lambda) - e^{-\frac{x_u}{\lambda}}(x_u + 2\lambda)\right), \qquad \mu_y = \lambda - \mu_w \quad (41)$$

Finally, the probability density of the final income $\underline{z}$ will be

$$f_z(z) = f_w(z - a\mu_y) \quad (42)$$

Now we apply the toy model assuming $\lambda = \mu = 1, x_0 = 0.2, x_u = 1.4, p_u = 0.4, a = 0.5$, resulting in $\mu_y = 0.32, a\mu_y = 0.16$. Figure 13 shows the variation of the tax, the income minus tax, and the redistributed income vs. the original income. Notice that for small income the tax is zero and thus the income minus tax equals the original income, while for large income it equals 60% of the original income (Figure 13, right).

Figure 14 depicts the probability density of the redistributed income in comparison to that of the original income. The differences are that in the redistributed income: (a) poverty below the level $a\mu_y = 0.16$ is eliminated; (b) the middle class, corresponding to incomes up to the mean, is more populated and amplified; (c) the rich lose income. As a result of these, both the entropy and the coefficient of variation have decrease in the redistributed income; the former from $\Phi = 1$ of the exponential distribution to $\Phi = 0.55$ (or Φ_μ from 1 to 0.73) and the latter from $\sigma/\mu = 1$ to $\sigma/\mu = 0.57$.

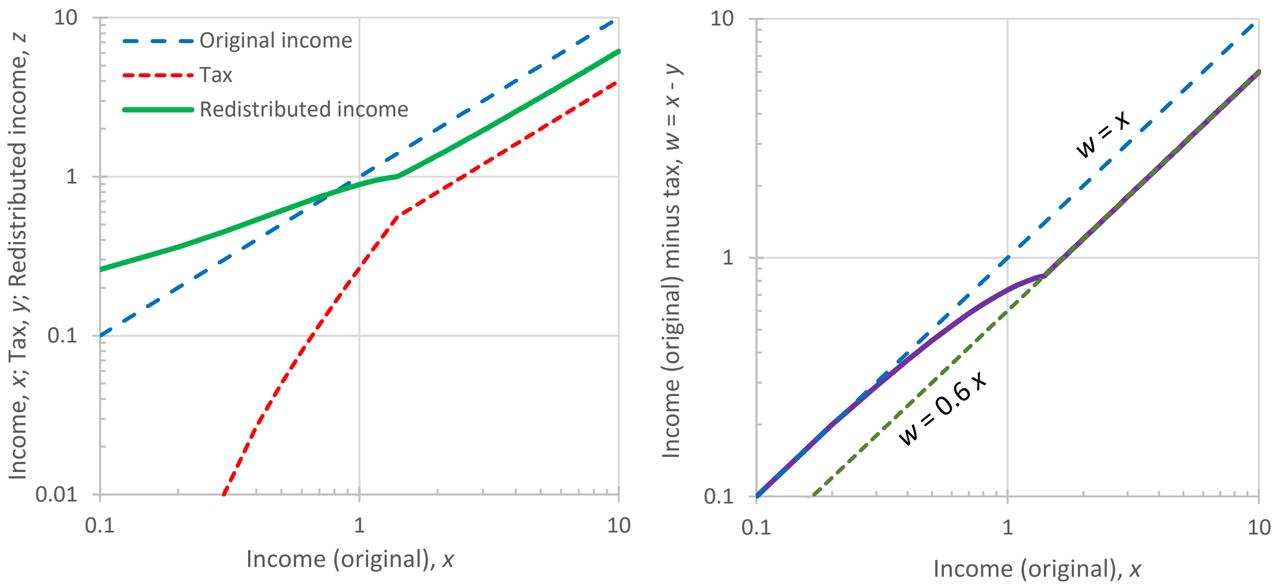


Figure 13. Variation of the indicated quantities with the original income (x) in the toy model: (**left**) tax and redistributed income; (**right**) original income minus tax.

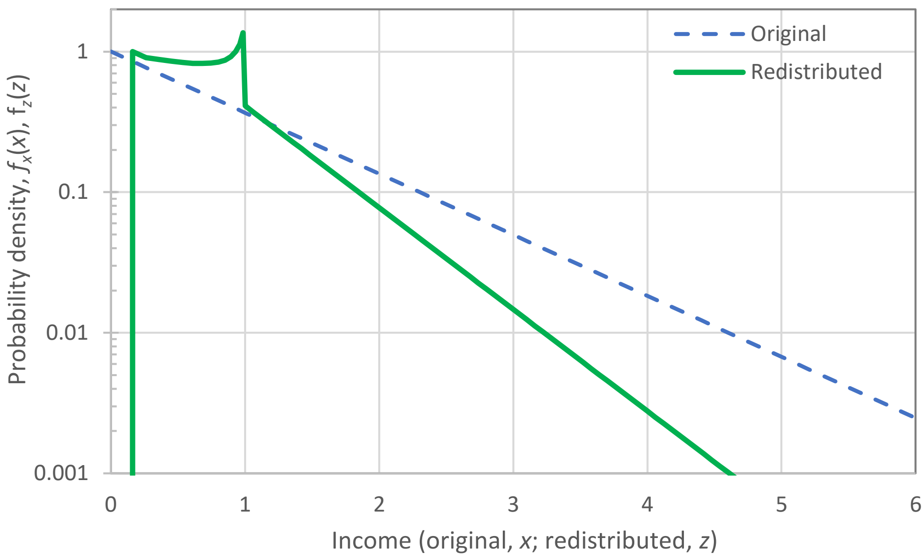


Figure 14. Comparison of the density function of the original income (x) and the final (redistributed) income (z) in the toy model.

4.4 Empirical investigation

The empirical investigation in this section provides comparison of the theoretical framework developed with real-world data. Sargentis et al. [106] have made similar comparisons, also intercomparing with the Lorenz curve [121-123] and the Gini coefficient [124-126], which are more standard measures of income distribution and socio-economic inequality [127-130]. In their comparisons, they used data given in tenths of the share of people from the lowest to the highest income versus share of income earned. The partitioning in tenths is a standard form of income data offered in relevant databases, yet it fully hides the behaviour of the tail, which, as we have seen, is extremely important for understanding structural characteristics of the economy and for quantifying inequality. We will provide additional evidence on this importance in this section.

For our purpose, we have searched for data given in higher resolution than tenths of people's income and we found such data for USA and Sweden. Even in this case, the information about the tail (the very rich people) is missing as the data values end at some level c with the last bunch of data given as "c and over". It is thus crucial to find a way to extrapolate the distribution function beyond c and estimate expectations based on this extrapolation.

It is consistent with our theoretical framework to assume that beyond c the following approximation is suitable:

$$f(x) = \mathrm{e}^{-x/k+b} \Rightarrow \overline{F}(x) = k\mathrm{e}^{b-\frac{x}{k}} = kf(x), \qquad x \geq c \tag{43}$$

where b and k are parameters to be estimated. The expectation of any function $g(\underline{x})$ can be calculated as

$$E[g(\underline{x})] \coloneqq \int_0^\infty g(x)f(x)\mathrm{d}x = A^g + B^g, \qquad A^g \coloneqq \int_0^c g(x)f(x)\mathrm{d}x\,, B^g \coloneqq \int_c^\infty g(x)f(x)\mathrm{d}x \tag{44}$$

The quantity A^g can directly be estimated from the available data, by approximating the integral with a sum. Assuming that the data are given in terms of the number of persons N_i with income between levels x_{i-1} and x_i, with $i = 1, \dots n$ and $x_n \equiv c$, we have:

$$\hat{A}^g = \sum_{i=1}^{n} g\left(\frac{x_{i-1} + x_i}{2}\right) \hat{f}_i\,(x_i - x_{i-1}), \qquad \hat{f}_i = \frac{N_i}{N(x_i - x_{i-1})}, \qquad N = \sum_{i=1}^{n} N_i + N_c \tag{45}$$

where N_c is the number of people with income $> c$.

The quantity B^g is estimated from the approximation (43). For the moment of order p of the distribution we have:

$$B_p \coloneqq \int_c^\infty x^p \mathrm{e}^{-ax+b}\mathrm{d}x = \frac{\mathrm{e}^b}{k^{1+p}}\Gamma\left(1 + p, \frac{c}{k}\right) \tag{46}$$

In particular for $p = 0{,}1{,}2$ we have

$$B_0 = k\mathrm{e}^{b-c/k}, \qquad B_1 = B_0(c + k), \qquad B_2 = B_0(k^2 + (c + k)^2) \tag{47}$$

Since both $f(c)$ and $\overline{F}(c)$ can directly be estimated from the data as $\hat{f}(c) \equiv \hat{f}_n$ and $\hat{\overline{F}}(c) = N_c/N$, we have

$$\hat{f}(c) = \mathrm{e}^{-c/k+b}, \qquad \hat{B}_0 = \hat{\overline{F}}(c) = k\hat{f}(c) \tag{48}$$

and by solving these equations we find the unknown parameters as

$$\hat{k} = \frac{\hat{\overline{F}}(c)}{\hat{f}(c)}, \qquad \hat{b} = \ln \hat{f}(c) + \frac{c}{\hat{k}} \tag{49}$$

This allows estimation of B^g for the expectation of any function $g(\underline{x})$, by replacing B_0, k and b with their estimates. In particular, for the entropy we have

$$\hat{B}_\Phi = \hat{B}_0\left(1 - \hat{b} + \frac{c}{\hat{k}}\right) \tag{50}$$

The data from the USA are available online thanks to the United States Census Bureau [131]; of those we choose to use the most recent available, those for year 2019 [132] and in particular those for the entire population irrespective of particular characteristics (sex, race, etc.). The empirical probability density and tail function (probability of exceedance), estimated from the data are shown in Figure 15, also in comparison with the entropy maximizing exponential distribution.

Here it is useful to remark that the detailed data cover only a small portion of the range of incomes, up to less than double the mean income. Thus, they provide little information on the distribution tail (the richest people). As a result, the extrapolation becomes very important. Without the extrapolation the mean income (i.e. the quantity A_1, according to the above notation) is 30 601 \$ and it becomes 53 336 \$ after the extrapolation (i.e.

after adding B_1). Note that the actual mean value, according to the source data [132] is 54 129 $, i.e. close to the extrapolated estimate, which suggests that the extrapolation model is not bad. Even more drastic is the change in the second moment: before extrapolation, it is 1.62 × 10^9 and after it 4.74 × 10^9, i.e. almost 3 times higher. The final (with extrapolation) estimate of the coefficient of variation is $\sigma/\mu = 1.11$, slightly higher than 1. The final estimate of entropy (for λ = 1 $) is $\Phi = 11.82$, and that of the standardized entropy $\Phi_\mu = 0.94$, slightly lower than 1.

Overall, the picture in Figure 15 suggests that the principle of maximum entropy with Lebesgue background measure seems to explain the income distribution. It is interesting that the frequency of mid-rich people, from the mean income to more than twice the mean, is somewhat overpredicted by the exponential distributions, and that of the very rich (with income more than thrice the mean) is underpredicted. The incomes of the poor and middle class do not differ from what is predicted by the principle of maximum entropy. Remarkably, the condition $\mu_{x\geq c}/\mu = 80\%$ is satisfied when $\overline{F}(c) = 42\%$, close to the value 43.9% of the exponential distribution (equation (33)) and substantially higher than 20%, thus suggesting the inappropriateness of the "80/20 rule".

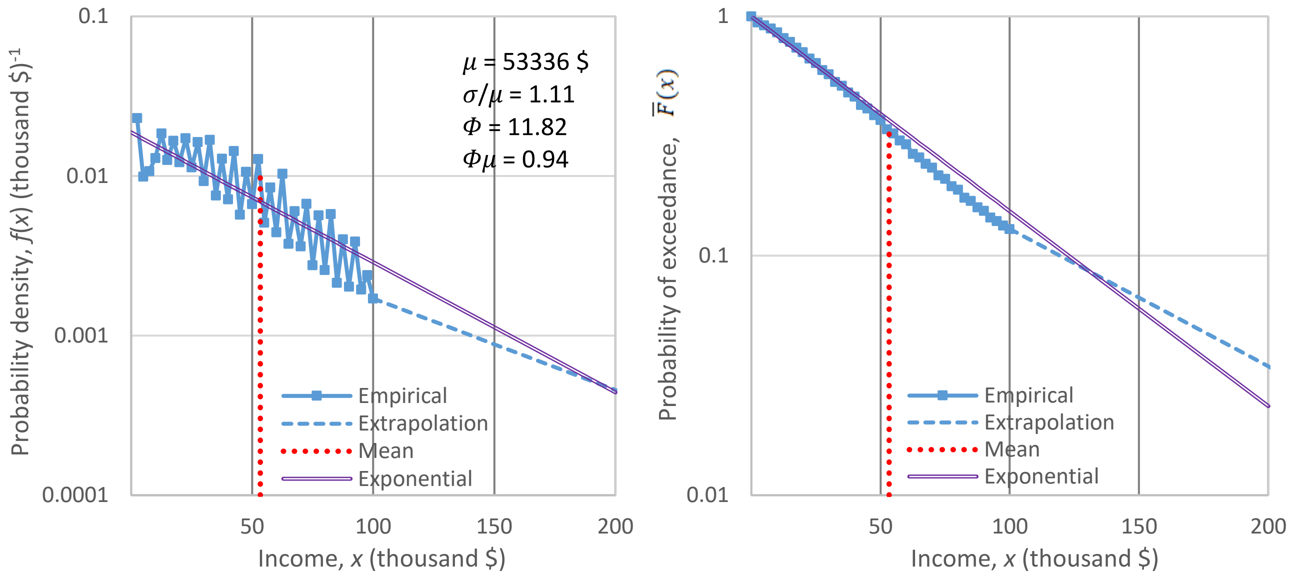
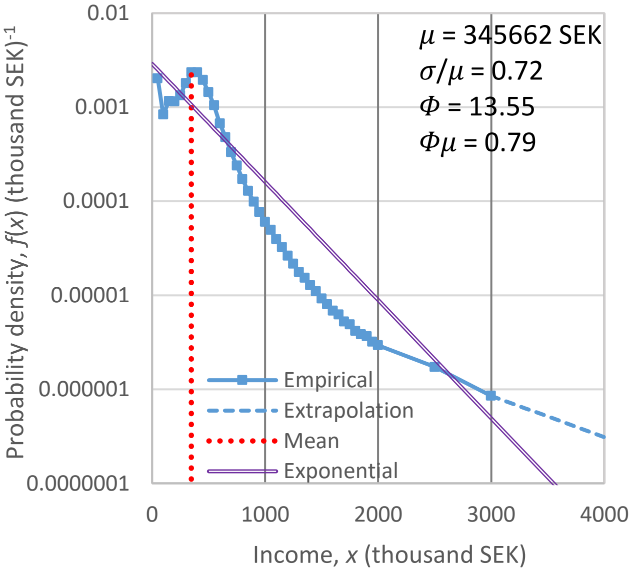


Figure 15. Illustration of the entropic framework using income data from the USA for year 2019; (**left**) probability density; (**right**) tail function (probability of exceedance).

Somewhat different is the picture of Sweden shown in Figure 16, again for the year 2019. The data, provided by Statistics Sweden [133], are more detailed than the USA data, covering a range of income about nine times the mean income [134]. The estimated statistics are also shown in the figure. Here, the graph is consistent with that of Figure 14 (our toy model) up to about five times the mean income, indicating the presence of a populated middle class and suggesting an effect of the taxing system. The standardized entropy ($\Phi_\mu = 0.94$) and the coefficient of variation ($\sigma/\mu = 0.72$) are lower than in the USA, suggesting lower inequality. Strikingly however, there is an opposite effect on the very rich, whose frequency is considerably higher than that predicted by the exponential distribution. A possible explanation would rely on globalization of the financial activities of the high-net-worth individuals. Again, the condition $\mu_{x\geq c}/\mu = 80\%$ is satisfied for $\overline{F}(c)$ substantially higher than 20%, namely for $\overline{F}(c) = 45\%$, close to the value 43.9% of the exponential distribution, thus suggesting again the inappropriateness of the "80/20 rule". Overall, the principle of maximum entropy again provides a good representation of the average behaviour.



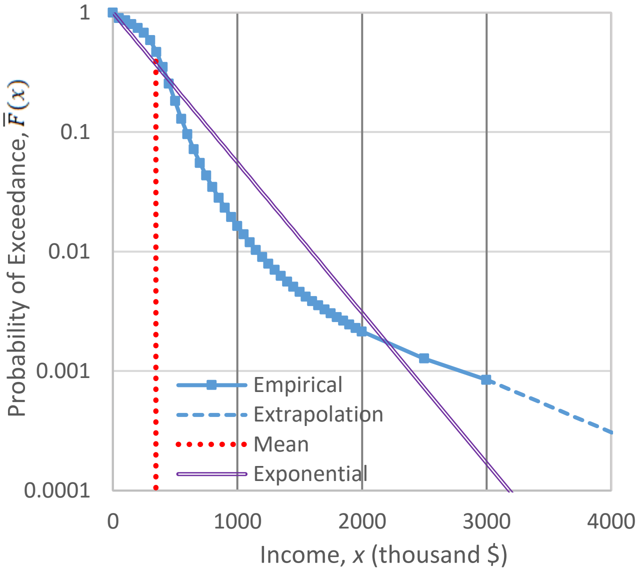


Figure 16. Illustration of the entropic framework using income data from Sweden for year 2019; (**left**) probability density; (**right**) tail function (probability of exceedance).

4. Discussion and conclusions

We have shown that the entropy is one of the most misunderstood concepts with rich and diverse interpretations which are continuously debated. With the transplantation of the scientific term to the colloquial language, the popular imagination has loaded "entropy" with almost every negative quality in the universe, in life and in society. For example, Thesaurus.com lists as its synonyms the words breakup, collapse, decay, decline, degeneration, destruction, worsening, falling apart, while in the site wordhippo.com the synonyms listed, in addition to those, amount to hundreds of words with negative meaning, including deterioration, chaos, havoc, confusion, disorder, disorganization, calamity etc. Also in scholarly articles, there is no shortage of negative associations, as quoted in the Introduction.

There are historical reasons, discussed in section 2, why the concept generated so many negative connotations. However, in the end of the 1940s, entropy acquired a clear and universal stochastic definition which is not a related to disorder. Furthermore, in the end of the 1950s, it was complemented by the principle of maximum entropy, which lies behind the Second Law, and gives explanatory and inferential power to the concept. By now, 60 years after, one would expect that entropy should have aborted the negative meanings, and be recognized as the driving force of natural change and the mother of creativity and evolution. This did not happen. Instead, it has been used as a spectre in social sciences, including economics and ecology, to promote neo-Malthusian ideas.

As social sciences are often contaminated by subjectivity and ideological influences, here we explore whether the maximum entropy, applied to economics and, in particular, to the distribution of wealth quantified by annual income, can give an objective description. We show that under plausible constraints related to the mean income, the principle of maximum entropy results in exponential distribution, bounded from above if we consider an upper technological limit, or unbounded otherwise. Historically, technology has played a major role in development and increase of the entropy of income. Under current conditions, technology no longer imposes a bounding condition on the economy, yet it remains an important factor in increasing wealth.

This entropy maximizing distribution emerges when the background measure has constant density, while if a hyperbolic background measure is used, the resulting distribution is Pareto. Based on real-world data, and in particular, those of the world's richest, in order to give a better idea on the distribution tail, we conclude that the exponential tail is not uncommon, while the Pareto tail appears particularly in anomalous periods.

Impressively, the latest period of pandemic resulted in unprecedent profits of the richest, with a clear Pareto tail.

We conclude that a constant (Lebesgue) density of the background measure is reasonable and that the entropy maximizing, under this measure, exponential distribution is connected to a stable economy. Furthermore, we view two different factors, both leading to reduction of entropy and modification of the stable exponential distribution, but in different directions. On the one hand, the organized societies use mechanisms of income redistribution in order to minimize poverty and enhance the middle class. On the other hand, an assumption can be made that politico-economic elites try to increase their profits, thus pointing toward a Pareto distribution, which populates more the poor and the very rich and reduces the middle class.

Using publicly available income data for USA and Sweden, we showed that the income distribution is consistent with the principle of the maximum entropy and in particular, with the exponential distribution. Yet the effect of the elites is visible as the distribution tails exceed that of the exponential. On the other hand, the data do not support the "80/20" rule, which is consistent with the Pareto distribution (with a specific value of the tail index). Specifically, the 80% of income is not generated by 20% of the population but by more than 40% thereof, in full consistence with the exponential distribution.

Overall, in this study we have tried to dispel the "bad name" of entropy in social sciences. We emphasized its connection with the plurality of options and we showed that increasing entropy is associated to increase of wealth. In addition, we showed that a standardized form of entropy can be used to quantify inequality.

We tried to make the paper self-contained and stand-alone, so that even a reader unfamiliar with entropy, with only basic knowledge on calculus and probability, could assimilate it. The mathematical framework we developed can readily be put to work on the simplest computational framework (e.g. a spreadsheet). The entire study is of exploratory character, as our priority was to show what we believe entropy really is, and under which conditions it could be applied to economics. Future work could open up additional options, thus increasing the entropy.

Supplementary Materials: N/A

Author Contributions: Conceptualization, DK; methodology, DK; software, DK; validation, DK and GFS; formal analysis, DK; investigation, DK and GFS; resources, DK and GFS; data curation, DK and GFS; writing—original draft preparation, DK and GFS; writing—review and editing, DK and GFS; visualization, DK and GFS; supervision, DK; project administration, N/A; funding acquisition, N/A. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding but was motivated by the scientific curiosity of the authors.

Data Availability Statement: The databases that have been used are referred to in detail in the citations given in the text and are publicly available.

Acknowledgments: A primitive version of this work by DK was presented in 2018 as a critique in the inauguration of the first volume of the T. Xanthopoulos's book trilogy "Requiem with Crescendo", a critique particularly focusing on the reasoning about the primitive "classless society". We thank him for inspiring this work and for discussions on relevant issues.

Conflicts of Interest: We declare no conflict of interest. We believe that anonymous reviewing may hide possible conflicts of interest and hence we favour eponymous reviewing.

References

1. Razumovskaya, V.A. Information entropy of literary text and its surmounting in understanding and translation, *Journal of Siberian Federal University: Humanities & Social Sciences 2* **2010**, 3 259-267. Available online: https://core.ac.uk/download/pdf/38633204.pdf (accessed on 9 August 2021).
2. Anger, S. Evolution and entropy. In *A Companion to British Literature, Scientific Contexts in the Nineteenth Century*, DeMaria, R., Chang, H.; Zacher, S., Eds. Wiley Online Library, 2014, doi: 10.1002/9781118827338.ch79.
3. Entropy poems. Available online: http://www.poemhunter.com/poems/entropy/ (accessed on 9 August 2021).
4. Hello poetry. Available online: http://hellopoetry.com/words/entropy/ (accessed on 9 August 2021).
5. Corpus of Contemporary American English. Available online: http://www.english-corpora.org/coca/ (accessed on 9 August 2021).
6. Word and Phrase Info. Available online: http://www.wordandphrase.info/frequencyList.asp (accessed on 9 August 2021).
7. Scopus database, Available online: https://www.scopus.com/ (accessed on 13 August 2021)
8. Overbury, R.E. Features of a closed-system economy. *Nature* **1973**, 242 (5400), 561-565.
9. Bailey, K.D. *Social Entropy Theory*, State University of New York, New York, USA, 1990.
10. Bailey, K.D. Social entropy theory: An overview. *Systems Practice* **1990**, 3(4), 365-382.
11. Mayer T.F. Review of "Social Entropy Theory" by Kenneth D. Bailey, *American Journal of Sociology* **1991**, 96 (6) 1544-1546, https://www.jstor.org/stable/2781914 (accessed on 13 August 2021).
12. Netto , V.M.; Meirelles, J.; Ribeiro, F.L. Social interaction and the city: the effect of space on the reduction of entropy, *Complexity* **2017,** vol. 2017, Article ID 6182503, doi: 10.1155/2017/6182503.
13. Dinga, E.; Tănăsescu, C.-R.; Ionescu, G.-M. Social entropy and normative network. *Entropy* **2020**, 22, 1051, doi: 10.3390/e22091051.
14. Balch, T. Hierarchic social entropy: An information theoretic measure of robot group diversity. *Autonomous Robots* **2000**, 8(3), 209-238.
15. Mavrofides, T.; Kameas, A.; Papageorgiou, D.; Los, A., On the entropy of social systems: a revision of the concepts of entropy and energy in the social context. *Syst. Res.* **2011**, 28, 353-368, doi: 10.1002/sres.1084.
16. Davis, P.J. Entropy and society: can the physical/mathematical notions of entropy be usefully imported into the social sphere? *Journal of Humanistic Mathematics* **2011**, 1 (1) 119-136, doi: 10.5642/jhummath.201101.09.
17. Saridis, G.N. Entropy as a philosophy. Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop, National Institute of Standards & Technology (NIST), Manufacturing Engineering Lab, Gaithersburg, MD, USA, 2004, http://apps.dtic.mil/sti/citations/ADA515701 (accessed on 15 August 2021).
18. Soddy, F. Wealth, *Virtual Wealth and Debt: The Solution of the Economic Paradox*, Britons Publishing Company, London 1933, Available online: http://www.fadedpage.com/showbook.php?pid=20140873 (accessed on 9 August 2021).
19. Mansel, D. Frederick Soddy: The scientist as prophet. *Annals of Science* **1992**, 49(4), 351-367, doi: 10.1080/00033799200200301.
20. Daly, H.E. The economic thought of Frederick Soddy. *History of Political Economy* **1980**, 12(4), 469–488. doi: 10.1215/00182702-12-4-469.
21. Georgescu-Roegen, N. *The Entropy Law and the Economic Process*; Harvard University Press, Cambridge, Mass. USA, 1971, doi: 10.4159/harvard.9780674281653.
22. Georgescu-Roegen, N. The entropy law and the economic process in retrospect. *Eastern Economic Journal* **1986**, 12, 3-25. Available online: https://www.jstor.org/stable/40357380 (accessed on 14 August 2021).
23. Beard, T.R. *Economics, Entropy and the Environment: The Extraordinary Economics of Nicholas Georgescu-Roegen*; Edward Elgar Publishing, London, UK, 1999.
24. Avery, J. S. Entropy and economics. In *Cadmus: The wealth of nations revisited, A papers series of the South-East European Division*, Vol I (4), 2012 Available online: http://cadmusjournal.org/node/146 (accessed on 9 August 2021).
25. Candeal, J.C.; De Miguel, J.R.; Induráin, E.; Mehta, G.B. Utility and entropy. *Economic Theory* **2001**, 17(1), 233-238, Available online: http://www.jstor.org/stable/25055372 (accessed on 31 July 2021).
26. Jakimowicz, A. The role of entropy in the development of economics. *Entropy* **2020**, 22(4), 452, doi: 10.3390/e22040452.
27. Jaynes, E.T. How should we use entropy in economics? University of Cambridge: Cambridge, UK. 1991. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.4606&rep=rep1&type=pdf (accessed on 9 August 2021).
28. Beard, T.R.; Lozada, G.A. *Economics, Entropy and the Environment*; Edward Elgar Publishing, London, UK, 1999.
29. Kümmel, R. *The Second Law of Economics: Energy, Entropy, and the Origins of Wealth*; Springer Science & Business Media, London, UK, 2011.
30. Faber, M.; Niemes, H.; Stephan, G. *Entropy, Environment and Resources: An Essay in Physico-Economics*; Springer-Verlag, London, UK, 1987.
31. Mayumi, K.; Giampietro, M. Entropy in ecological economics. In: *Modelling in Ecological Economics*, Proops, J.; Safonov, P. Eds. 80-101, Edward Elgar, Cheltenham, UK, 2004. Available online: https://books.google.gr/books?id=ar4rEAAAQBAJ (accessed on 10 August 2021).
32. Fisk, D. Thermodynamics on Main Street: When entropy really counts in economics. *Ecological Economics* **2011**, 70(11), 1931-1936, doi: 10.1016/j.ecolecon.2011.06.012.
33. Lozada, G.A. Entropy, free energy, work, and other thermodynamic variables in economics. *Ecological Economics* **2006**, 56(1), 71-78, doi: 10.1016/j.ecolecon.2004.12.019.

34. Giffin, A., From physics to economics: An econometric example using maximum relative entropy, *Physica A: Statistical Mechanics and its Applications* **2009**, 388(8), 1610-1620, doi: 10.1016/j.physa.2008.12.066.
35. Townsend, K.N. Is the entropy law relevant to the economics of natural resource scarcity? Comment. *Journal of Environmental Economics and Management* **1992**, 23(1), 96-100, doi: 10.1016/0095-0696(92)90044-W.
36. Burkett, P. Entropy in ecological economics: a Marxist intervention. In *Marxism and Ecological Economics*, Leiden, The Netherlands, 2006, doi: 10.1163/9789047408567_007
37. Ruth, M. Entropy, economics, and policy. In *Thermodynamics and The Destruction of Resources*; Bakshi, B.; Gutowski, T.; Sekulić, D. Eds.; Cambridge University Press, Cambridge, UK, 2011, 402-422, doi: 10.1017/CBO9780511976049.
38. Raine, A.; Foster, J.; Potts, J. The new entropy law and the economic process, *Ecological Complexity* **2006**, 3(4), 354-360, doi: 10.1016/j.ecocom.2007.02.009.
39. McMahon, G.F.; Mrozek, J.R. Economics, entropy and sustainability, *Hydrological Sciences Journal* **1997**, 42(4), 501-512, doi: 10.1080/02626669709492050.
40. Smith, C.E.; Smith, J.W. Economics, ecology and entropy: The second law of thermodynamics and the limits to growth. *Population and Environment* **1996**, 17(4), 309–321, doi: 10.1007/BF02208231.
41. Hammond, G.P.; Winnett, A.B. The influence of thermodynamic ideas on ecological economics: an interdisciplinary critique. *Sustainability* **2009**, 1, 1195-1225, doi: 10.3390/su1041195.
42. Kovalev, A.V. Misuse of thermodynamic entropy in economics. *Energy* **2016**, 100, 129-136, doi: 10.1016/j.energy.2016.01.071.
43. Scharfenaker, E.; Yang, J. Maximum entropy economics: where do we stand? *Eur. Phys. J.* **2020**, Spec. Top. 229, 1573–1575, doi: 10.1140/epjst/e2020-000030-3.
44 Scharfenaker, E.; Yang, J. Maximum entropy economics. *Eur. Phys. J.* **2020**, Spec. Top. 229, 1577–1590, doi: 10.1140/epjst/e2020-000029-4.
45. Ryu, H.K. Maximum Entropy estimation of income distributions from Bonferroni indices. In *Modeling Income Distributions and Lorenz Curves. Economic Studies in Equality, Social Exclusion and Well-Being*, Chotikapanich D. Ed.; vol 5. Springer, New York, NY. 2008, doi: 10.1007/978-0-387-72796-7_10
46. Fu, Q.; Villas-Boas, S.B.; Judge G. Entropy-based China income distributions and inequality measures, *China Economic Journal* **2019**, 12(3), 352-368, doi: 10.1080/17538963.2019.1570620.
47. Mayer, A.L.; Donovan, R.P.; Pawlowski, C.W. Information and entropy theory for the sustainability of coupled human and natural systems. *Ecology and Society* **2014**, 19(3), 11, doi: 10.5751/ES-06626-190311.
48. Liddell, Scott, Jones (LSJ): Ancient Greek Lexicon, Available online: https://lsj.gr/wiki/ἐντροπία (accessed on 31 July 2021).
49. Clausius, R. A contribution to the history of the mechanical theory of heat. *Phil. Mag.* **1872**, 43, 106–115.
50. Clausius, R. *The Mechanical Theory of Heat: With Its Applications to the Steam-Engine and to the Physical Properties of Bodies*. J. van Voorst, London, UK, 1867, Available online: http://books.google.gr/books?id=8LIEAAAAYAAJ (accessed on 31 July 2021).
51. Clausius, R. Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physik und Chemie* **1865**, 125 (7), 353–400, Available online: http://books.google.gr/books?id=qAg4AAAAMAAJ (accessed on 31 July 2021).
52. Hill, G.; Holman, J. *Entropy – the Driving Force of Change*, Royal Society of Chemistry, London, ISBN: 0-15186-967-X, 17 pp., 1986.
53. Atkins, P. *Galileo's Finger: The Ten Great Ideas of Science*. Oxford University Press, New York, USA, 2003.
54. Atkins, P. 2007. Four Laws that Drive the Universe, Oxford University Press, Oxford, UK, 131 pp.
55. Boltzmann, L. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. *Wien. Ber.* **1877**, 76, 373-435.
56. Swendsen, R.H. Statistical mechanics of colloids and Boltzmann's definition of the entropy. *American Journal of Physics* **2006**, 74(3), 187-190.
57. Gibbs, J.W. *Elementary Principles of Statistical Mechanics*; Yale University Press, New Haven, USA, 1902, 244 pp. (Reprinted by Dover, New York, 1960), Available online: https://www.gutenberg.org/ebooks/50992 (accessed on 31 July 2021).
58. Shannon, C.E. The mathematical theory of communication. *Bell System Technical Journal* **1948**, 27 (3), 379-423.
59. Robertson, H.S. *Statistical Thermophysics*; Prentice Hall, Englewood Cliffs, NJ, USA, 1993, 582 pp.
60. Brissaud, J.-B. The meanings of entropy. *Entropy* **2005**, 7, 68-96, doi: 10.3390/e7010068.
61. Koutsoyiannis, D. Hurst-Kolmogorov dynamics as a result of extremal entropy production. *Physica A* **2011**, 390(8), 1424–1432, doi: 10.1016/j.physa.2010.12.035.
62. Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine*; MIT Press, Cambridge, Mass., USA, 1948, 212 pp.
63. von Neumann, J. Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata Studies* **1956**, 43-98, doi: 10.1515/9781400882618-003.
64. Szilard, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik* **1929**, 53(11), 840-856.
65. Jaynes, E.T. Information theory and statistical mechanics. *Physical Review* **1957**, 106(4), 620-630.
66. Swendsen, R.H., 2011. How physicists disagree on the meaning of entropy. *Am. J. Phys.*, 79 (4), 342-348.
67. Koutsoyiannis, D. Physics of uncertainty, the Gibbs paradox and indistinguishable particles. *Studies in History and Philosophy of Modern Physics* **2013**, 44, 480–489, doi: 10.1016/j.shpsb.2013.08.007.
68. Koutsoyiannis, D. Entropy: from thermodynamics to hydrology. *Entropy* **2014**, 16(3), 1287–1314, doi: 10.3390/e16031287.

69. Boltzmann, L. On the indispensability of atomism in natural science. First published in *Annalen der Physik und Chemie* **1897**, 60, 231 (republished: Theoretical Physics and Philosophical Problems, 41-53, Springer, Dordrecht, 1974, doi: 10.1007/978-94-010-2091-6_5.
70. Boltzmann, L. On the necessity of atomic theories in physics. *The Monist* **1901**, 12 (1), 65-79, Available online: https://www.jstor.org/stable/27899285 (accessed on 31 July 2021).
71. Boltzmann, L. *Lectures on Gas Theory*, Dover Publications, New York, USA, 1995, 443 pp.
72. Darrow K.K. The concept of entropy, *Am. J. Phys.* **1944**, 12, 183, doi: 10.1119/1.1990592.
73. Wiener, N. Time, communication, and the nervous system. *Annals of the New York Academy of Sciences* **1948**, 50(4), 197-220.
74. Chaitin, G.J. Computers, paradoxes and the foundations of mathematics. *American Scientist* **2002**, 90(2), 164-171.
75. Bailey, K.D. Entropy Systems Theory: Systems Science and Cybernetics, Eolss Publishers, Oxford, UK, 2009, 169 pp.
76. Koutsoyiannis, D. A random walk on water, *Hydrology and Earth System Sciences* **2010**, 14, 585–601, doi: 10.5194/hess-14-585-2010.
77. Google Scholar, https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=author%3Ah-kissinger+%22world+order%22 (accessed on 8August 2021).
78. Google Scholar, https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=author%3Ah-kissinger+%22new+world+order%22 (accessed on 8 August 2021).
79. Kissinger, H.A., The chance for a new world order. *The New York Times*, 12 Jan. 2009, https://www.nytimes.com/2009/01/12/opinion/12iht-edkissinger.1.19281915.html (accessed on 8 August 2021).
80. Kissinger, H.A. The coronavirus pandemic will forever alter the world order, *The Wall Street Journal*, 3 April 2020, https://www.wsj.com/articles/the-coronavirus-pandemic-will-forever-alter-the-world-order-11585953005 (accessed on 8 August 2021).
81. Onuf P.S.; Onuf, N.G. American constitutionalism and the emergence of a liberal world order. In: *American Constitutionalism Abroad: Selected Essays in Comparative Constitutional History*; Billias, G.A. Ed. Greenwood Press, New York, USA, 1990.
82. Schwab, K.; Malleret, T. *Covid-19: The Great Reset*; World Economic Forum, Geneva, Switzerland, 2020.
83. Google Books Ngram Viewer https://books.google.com/ngrams/ (accessed on 1 August 2021).
84. Michel, J.B.; Shen, Y.K.; Aiden, A.P.; Veres, A.; Gray, M.K.; Pickett, J.P.; Hoiberg, D.; Clancy, D.; Norvig, P.; Orwant, J.; Pinker, S. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* **2011**, 331(6014), 176-182.
85. Wright, P.G. Entropy and disorder. *Contemporary Physics* **1970**, 11(6), 581-588, doi: 10.1080/00107517008202196.
86. Leff, H.S. Removing the mystery of entropy and thermodynamics—Part V. *Phys. Teach.* **2012**, 50(5), 274-276, doi: 10.1119/1.3703541.
87. Styer, D. Entropy as disorder: history of a misconception, *Phys. Teach.* **2019**, 57, 454, doi: 10.1119/1.5126822.
88. Adams, H. *The Education of Henry Adams*, Houghton Mifflin Company, Boston, USA, 1918.
89. Nordangård, J. Rockefeller: Controlling the Game, Stiftelsen Pharos, Lidingö, Sweden, 2019, 386 pp., ISBN: 978-91-985843-0-1.
90. van der Sluijs, J. Uncertainty as a monster in the science–policy interface: Four coping strategies. *Water Science and Technology* **2005**, 52(6), 87-92.
91. Curry, J.A.; Webster, P.J. Climate science and the uncertainty monster. *Bulletin of the American Meteorological Society* **2011**, 92(12), 1667-1682.
92. Koutsoyiannis D.; Makropoulos, C.; Langousis, A.; Baki, S.; Efstratiadis, A.; Christofides, C.; Karavokiros, G.; Mamassis, N. Interactive comment on "HESS Opinions, Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability" by D. Koutsoyiannis et al. *Hydrol. Earth Syst. Sci. Discuss.* **2008**, 5, S1761–S1774, www.hydrol-earth-syst-sci-discuss.net/5/S1761/2008/ (accessed on 14 August 2021).
93. Marc Chagall, Palette (1974), The Metropolitan Museum of Art, New York, USA, https://www.metmuseum.org/art/collection/search/493731 (accessed on 13 August 2021).
94. Marc Chagall, Self-Portrait with Seven Fingers, 1913, https://en.wikipedia.org/wiki/Self-Portrait_with_Seven_Fingers (accessed 13 August 2021).
95. Guye, C.-E. L'évolution Physico-chimique. E. Chiron, Paris, 1922.
96. Brillouin, L. Life, thermodynamics, and cybernetics. In *Maxwells Demon. Entropy, Information, Computing*; Princeton University Press, Princeton, New Jersey, USA, 89-103, 1949, doi: 10.1515/9781400861521.
97. Brillouin, L., Thermodynamics and information theory. *American Scientist* **1950**, 38 (4), 594-599, https://www.jstor.org/stable/27826339 (accessed on 1 August 2021).
98. Schrödinger, E. *What is Life? The Physical Aspect of the Living Cell*; Cambridge University Press, Cambridge, UK, 1944.
99. Lago-Fernández, L.F.; Corbacho, F. Using the negentropy increment to determine the number of clusters. In I*nternational Work-Conference on Artificial Neural Networks*, Springer, Berlin, Germany, 2009, pp. 448-455.
100. Larouche Jr, L.H., On Larouche's Discovery, 1993, http://www.archive.schillerinstitute.com/fidelio_archive/1994/fidv03n01-1994Sp/fidv03n01-1994Sp_037-on_larouches_discovery-lar.pdf (accessed on 1 August 2021).
101. Kolmogorov, A.N. Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Math.* **1933,** 2.
102. Kolmogorov, A.N., *Foundations of the Theory of Probability*, 2nd English Edition, 84 pp. Chelsea Publishing Company, New York, 1956.
103. Jaynes, E.T., 2003. *Probability Theory: The Logic of Science*, Cambridge Univ. Press, Cambridge, UK, 728 pp.
104. Uffink, J. Can the maximum entropy principle be explained as a consistency requirement? *Studies In History and Philosophy of Modern Physics* **1995**, 26 (3), 223-261.

105. Koutsoyiannis, D. *Stochastics of Hydroclimatic Extremes - A Cool Look at Risk*, Kallipos, Athens, Greece, 2021, ISBN: 978-618-85370-0-2, 333 pp.http://www.itia.ntua.gr/2000/ (accessed on 1 August 2021).
106. Sargentis, G.-F.; Iliopoulou, T.; Dimitriadis, P.; Mamassis, N.; Koutsoyiannis, D. Stratification: An Entropic View of Society's Structure. *World* **2021**, *2*, 153-174, doi:10.3390/world2020011.
107. Pareto V. La courbe de la répartition de la richesse (The curve of the distribution of wealth). In Pareto, V., 1967. *Ecrits sur la courbe de la répartition de la richesse*. Librairie Droz, Geneva, Switzerland, pp. 1–15.
108. Mornati F. (2018) The Law of Income Distribution and Various Statistical Complements. In: *Vilfredo Pareto: An Intellectual Biography Volume II*; Palgrave Studies in the History of Economic Thought, Palgrave Macmillan, Cham., UK, doi: 10.1007/978-3-030-04540-1_10.
109. Pareto V. Manual of Political Economy, Oxford University Press, New York, USA, 2014.
110. Blanchet, T.; Fournier, J.; Piketty T. Generalized Pareto Curves: Theory and Applications, WID.world Working Paper, 2017/3, Available online: https://wid.world/document/blanchet-t-fournier-j-piketty-t-generalized-pareto-curves-theory-applications-2017/ (accessed on 7 February 2021).
111. Craft, R.C.; Leake, C. The Pareto principle in organizational decision making, *Management Decision* **2002**, 40(8), 729-733, doi: 10.1108/00251740210437699.
112. Juran, J.M. The non-Pareto principle; mea culpa; *The Archives of Juran Institute*, 1975, Available online: https://www.juran.com/wp-content/uploads/2021/03/The-Non-Pareto-Principle-1974.pdf (accessed on 16 August 2021).
113. Juran, J.M. Universals in management planning and controlling. *Management Review* **1954**, 43(11), 748-761, Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.1456&rep=rep1&type=pdf (accessed on 16 August 2021).
114. The Complete World Billionaire Lists. Available online: https://stats.areppim.com/stats/links_billionairexlists.htm (accessed on 13 July 2021).
115. WaybackMachine. Available online: https://archive.org/web/, search field "www.forbes.com/billionaires/" (accessed on 13 July 2021).
116. WaybackMachine. Available online: https://archive.org/web/, search field "https://www.bloomberg.com//billionaires/" (accessed on 13 July 2021).
117. Marx, K.; Engels, F. *The German Ideology* (Vol. 1); International Publishers, New York, USA, 1970, Available online: https://archive.org/details/germanideology00marx/ (accessed on 11 August 2021).
118. Selsam, H.; Martel, H. *Reader in Marxist Philosophy: From the Writings of Marx, Engels and Lenin*; Rabul Foundation, Lucknow, India, 2010, 380 pp.
119. Koutsoyiannis, D. The political origin of the climate change agenda, Self-organized lecture, doi: 10.13140/RG.2.2.10223.05283, School of Civil Engineering – National Technical University of Athens, Athens, Greece, 2020.
120. Koutsoyiannis, D.; Mamassis, N. From mythology to science: the development of scientific hydrological concepts in the Greek antiquity and its relevance to modern hydrology. *Hydrology and Earth System Sciences* **2021**, 25, 2419–2444, doi: 10.5194/hess-25-2419-2021.
121. Tresch, R.W. The Social Welfare Function in Policy Analysis. In *Public Finance*, 3rd Ed., Ch. 4, Tresch R.W. Ed.; Academic Press, 2015, 57-78, ISBN 9780124158344, doi: 10.1016/B978-0-12-415834-4.00004-2.
122. Lorenz Curve. Available online: http://www.sciencedirect.com/topics/economics-econometrics-and-finance/lorenz-curve (accessed on 5 October 2020).
123. Bellù, L. G.; Liberati, P. Charting Income Inequality: The Lorenz Curve; Food and Agriculture Organization of the United Nations (FAO), 2005. Available online: http://www.fao.org/3/a-am391e.pdf (accessed on 28 December 2020)
124. Income inequality. Available online: https://ourworldindata.org/income-inequality (accessed on 5 October 2020).
125. Gini Index. Available online: https://www.investopedia.com/terms/g/gini-index.asp (accessed on 5 October 2020).
126. Bellù, L.G.; Liberati, P. Inequality Analysis. The Gini Index; Food and Agriculture Organization of the United Nations (FAO), 2006. Available online: http://www.fao.org/3/a-am352e.pdf (accessed on 28 December 2020).
127. Atkinson, A.B.; Bourguignon, F. Income Distribution. In International *Encyclopedia of the Social & Behavioral Sciences*; Smelser, N.J.; Baltes, P.B., Eds., Pergamon, 2001, 7265-7271, ISBN 9780080430768, doi: 10.1016/B0-08-043076-7/02236-1.
128. Morelli, S.; Smeeding, T.; Thompson, J. Post-1970 Trends in within-country inequality and poverty: rich and middle-income countries. In *Handbook of Income Distribution*; Atkinson, A.B.; Bourguignon, F., Eds., Elsevier, Ch. 8 Vol. 2, 2015, pp. 593-696, doi: 10.1016/B978-0-444-59428-0.00009-6.
129. Förster, M. F.; Tóth, I. G. Cross-country evidence of the multiple causes of inequality changes in the OECD area. In *Handbook of Income Distribution*; Atkinson, A.B.; Bourguignon, F., Eds., Elsevier, Ch. 19 Vol. 2, 2015, pp. 1729-1843, doi: 10.1016/B978-0-444-59429-7.00020-0.
130. Income Distribution, Redefining Capitalism in Global Economic Development, 2017, Available online: http://www.sciencedirect.com/topics/economics-econometrics-and-finance/income-distribution (accessed on 5 October 2020).
131. United States Census Bureau, PINC-08. Source of Income-People 15 Years Old and Over, by Income of Specified Type, Age, Race, Hispanic Origin, and Sex, Available online: https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-08.2019.html (accessed on 29 July 2021).
132. United States Census Bureau, PINC-08. Source of Income in 2019--People 15 Years Old and Over by Income of Specified Type in 2019, Age, Race, Hispanic Origin, and Sex, Available online: https://www2.census.gov/programs-surveys/cps/tables/pinc-08/2020/pinc08_1_1_1.xlsx (accessed on 29 July 2021).

133. SCB – Statistics Sweden, Income and tax statistics, Available online: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/household-finances/income-and-income-distribution/income-and-tax-statistics/ (accessed on 29 July 2021).
134. SCB – Statistics Sweden, Number of persons in brackets of total earned income by age 2019, Available online: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/household-finances/income-and-income-distribution/income-and-tax-statistics/pong/tables-and-graphs/income--persons-the-entire-country/total-income-from-employment-and-business-2019--by-income-brackets/ (accessed on 29 July 2021).