Entropy and infinite noise

About self and double references

Matthias Olschowy

April 24, 2022

Version 1.10

This is private research financed from my own funds. Therefore I am grateful for any donation. Payment information is available *here*.

Contents

Foreword

I would like to describe here in the preface the motives that led me to deal with entropy and complexity, to understand them and finally to write this book. First of all, there is the shadowy existence of complexity in physics, which has been spoiled by success for centuries, and by which it is not perceived, which is to a good part due to one of the characteristics of complexity itself, because it sets limits, but no hard ones. Thus, it does not impose itself, and the perception that it exists and is accompanied by a limitation is therefore by no means compelling. The lack of perception, however, is also due to the fact that entropy, which is a physical concept, is still not understood in physics, although it was introduced as a concept and physical state quantity by Clausius 170 years ago. On the other hand, its understanding is indispensable for the study of large systems, since it is an unavoidable companion of every large system. Thus, it is a part of reality and the study of it is a turn towards reality.

The essence of complexity is revealed much more clearly in economics. It was the place where I started my first thoughts about complexity. To be concrete, after my studies in theoretical physics, I was an employee of a company which, among many other things, was developing an international personnel accounting system. Very soon, I asked myself the simple question of why these systems actually have to be extremely complicated. The answer is just as simple. A software program, and especially a payroll system, is developed with the goal of being able to process many individual cases together. The essence of this set of individual cases is that they all differ at least a little from each other with the consequence that the program must necessarily reflect this diversity. And this is the simple essence of complex systems, namely their sheer size, which is in each case a consequence of the multiplicity of somehow distinguishable parts of a whole or parts between which a distinction must be drawn.

Another reason for writing this book is the importance of complexity for the conditions of human existence. Here I perceive a parallelism with physics, since according to my observation there is no sufficient awareness of the omnipresence of complexity in human existence, which also sets limits here, so that not everything is feasible, and if it is, then with increased effort, which we call work. We unconsciously perceive the complexity of a big city when we retreat to the silence of a hotel room. The essence of this retreat is diminishment, for the room with its walls presents a smaller world. Those who build a cabin in the wilderness do so for the same reason, to escape the complexity of nature. More precisely, however, it is not only nature in general and its complexity, but already that of the human body and subsequently that of its interaction with the outside. Thus, the history of man is the attempt to reduce the complexity of his own being and to limit his interaction with nature, i.e. simplify and control them.

Finally, the social and political context is dominated by complexity, which - here the dark side of complexity reveals itself - pervades the unconscious like a monster and dictates thinking as well as action. It is the breeding ground on which ideologies grow. There are two ways to reduce complexity. One has already been mentioned. It consists of keeping the system small by limiting it. The other is to avoid and reduce differences. Thus, every society is more or less uniformed, by culture, customs and laws, in which by no means only, but essentially also sameness manifests itself. It suggests itself that societies which are free in principle increasingly fall into a schism through this dualism. Here, one might say, wisdom is called for, more than ever. In this context, justice in the sense of equality, however understood, is the place where opinions differ. The problem now and the tragedy of justice is related to complexity, to order and disorder. For on the one hand, justice is only possible in an orderly system, since in the disorder of anarchy the unbridled right of the strongest applies, whereby inequality and, as a consequence, necessarily injustice are established. At the same time, order is only possible through distinction and thus inequality. Order is therefore on the one hand the precondition of justice and on the other hand its counterpart. Here, too, it is to be hoped that an awareness of the contradiction to which justice is subject will awaken. There is no perfect solution, only wisdom.

Overall, I hope to have made clear that complexity, order, and disorder are inherent features of human existence and the universe. Ultimately, I wrote the book to raise awareness of the ubiquity of complexity.

Chapter 1

Introduction

As already stated in the preface, in my opinion the meaning of entropy in physics is still not understood today, although it was already introduced by *Rudolf Clausius* in 1854.[1] First of all, the anecdote of a conversation in 1939 between *John von Neumann* and *Claude Shannon*, which is also reproduced in the said article of Wikipedia, gives reason for this:

> *My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons: In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.*[6]

Here, let us just say that *John von Neumann*, being one of the most authoritative thinkers on entropy at that time, used the formulation[2]

$$S = -k_B \sum_i p_i \log(p_i) \tag{1.1}$$

proposed by *J. Williard Gibbs* in the context of classical thermodynamics and transferred it over to quantum mechanics. Now, on the one hand, this formulation is regarded as a generalization of the *Boltzmann* entropy

$$S = -k_B \log(W) \tag{1.2}$$

as can be seen, for example, from the following two quotations:[3] [4]

> *The microstates of such a thermodynamic system are not equally probable—for example, high energy microstates are less probable than low energy microstates for a thermodynamic system kept at a fixed temperature by allowing contact with a heat bath. For thermodynamic systems where microstates of the system may not have equal probabilities, the appropriate generalization, called the Gibbs entropy...*[3]

---

[1] https://en.wikipedia.org/wiki/History_of_entropy

[2] https://en.wikipedia.org/wiki/Entropy_(statistical_thermodynamics)#Gibbs_entropy_formula

[3] https://en.wikipedia.org/wiki/Boltzmann's_entropy_formula

[4] https://en.wikipedia.org/wiki/Entropy_(statistical_thermodynamics)

> *Boltzmann's entropy describes the system when all the accessible microstates are equally likely. It is the configuration corresponding to the maximum of entropy at equilibrium. The randomness or disorder is maximal, and so is the lack of distinction (or information) of each microstate.*[4]

On the other hand, the abstract of the June 2019 article *Gibbs and Boltzmann Entropy in Classical and Quantum Mechanics* points to the opposite:[5]

> *The two approaches do not, however, necessarily agree for non-equilibrium systems. For those, we argue that the Boltzmann entropy is the one that corresponds to thermodynamic entropy, in particular in connection with the second law of thermodynamics.*[1]

The conclusion of the same article:

> *The Gibbs entropy (1) is an efficient tool for computing entropy values in thermal equilibrium when applied to the Gibbsian equilibrium ensembles ρ, but the fundamental definition of entropy is the Boltzmann entropy (2).*[1]

In this context, Boltzmann's momentous thesis should be mentioned, according to which the W in his formula 1.2 is equal to the number of states in the state space, which is shown by the following quote entitled *Boltzmann's principle*[6]:

> *Ludwig Boltzmann defined entropy as a measure of the number of possible microscopic states (microstates) of a system in thermodynamic equilibrium, consistent with its macroscopic thermodynamic properties, which constitute the macrostate of the system.*[4]

According to the quotation, *Boltzmann* limited the principle to the state of thermodynamic equilibrium. Nevertheless, it is considered valid also for arbitrary states. While the thermodynamic equilibrium is linked in theory with the uniform distribution, the idea is that the general state and thus also the thermodynamic disequilibrium are described by general distributions whose relative frequencies arise by counting states. The counting is preceded by a state classification, for instance by the energy of the system. The states within a class are counted, e.g. those with a certain energy or within an energy interval $[E, E + dE]$. The entropy of the system is defined in this conception then by the Gibbs entropy 1.1 of the counted relative frequencies, which expresses the following quotation from Wikipedia to the Gibbs entropy formula:

> *The macroscopic state of a system is characterized by a distribution on the microstates. The entropy of this distribution is given by the Gibbs entropy formula, named after J. Willard Gibbs. For a classical system (i.e., a collection of classical particles) with a discrete set of microstates, if E_i is the energy of microstate i, and p_i is the probability that it occurs during the system's fluctuations, then the entropy of the system is ... This definition remains meaningful even when the system is far away from equilibrium. Other definitions assume that the system is in thermal equilibrium, either as an isolated system, or as a system in exchange with its surroundings.*[4]

Definition of counting principle: I would like to call this generalized principle described above the counting principle. Its meaning is firstly the definition of entropy for a physical system on the basis of counting and equations 1.1 or 1.2. Secondly, a product rule belongs to the counting principle, according to which – in the logic of the Cartesian product[7] – the number of states of a total system decomposed into independent subsystems is equal to the product of the

---

[5] https://arxiv.org/pdf/1903.11870.pdf

[6] https://en.wikipedia.org/wiki/Entropy_(statistical_thermodynamics)#Boltzmann's_principle

[7] https://en.wikipedia.org/wiki/Cartesian_product

subsystem's numbers of states. From the product rule in turn follows the additivity and thus the extensivity of entropy for physically independent subsystems. For example, if two systems are equally distributed and have state numbers W_1 and W_2, then it follows from their independence and, in this case, from the validity of the product rule that the total system has $W_1 \cdot W_2$ states, so that the entropy is additive because of $\log(W_1) + \log(W_2) = \log(W_1 \cdot W2_)$.[8]

Regardless of the counting principle, the passages quoted above show that even today, almost 170 years after the discovery of thermodynamic entropy, there are contradictory views on what is to be understood by it. The counting principle, however, is the common basis of both views and until today the pivot for the understanding of entropy, to which the development from classical mechanics to quantum mechanics has an essential part, because in classical mechanics already the classification of particle states in the continuum of the position space and the momentum space is not possible or in the case of the discretization of the continuous space simply undetermined. In quantum mechanics, however, the energy spectrum of a system is already discrete, so that at least the classification is unambiguous.

According to what has been said, the counting principle together with the product rule is a fundamental part of the entropy definition. In the following, however, I would like to present a simple model for which the product rule with its meaning for the extensivity definitely has no validity:

For the model, we restrict ourselves to particles in the spatial domain. More precisely, we consider the states not of individual particles, but of particle collectives.[9] Let the system be divided into subsystems, i.e. into cells $C_1, C_2, \cdots$ of finite size in space, and let N particles be distributed among the cells, and let them be indistinguishable, but not the cells. By being distributed among the cells and thus assigned to them, the particles inherit the distinctness of the cells. Within a cell we then have a collective of indistinguishable particles, while those of different cells are distinguishable. For this model we now calculate the Boltzmann entropy for a system on the basis of the counting principle in analogy to the calculation of the entropy of mixing[10] and obtain – except for the factor of the particle number N – formally the Gibbs entropy, as will be shown shortly. To be concrete, let the distribution of particles among cells $C_1, C_2, \cdots$ be determined by the frequencies $n_1, n_2, \cdots$ If all N particles are distinguishable, then there are $N!$ permutations of the particles corresponding to $N!$ many distinguishable states. Instead, because the n_i particles within each C_i are indistinguishable, the total number of distinguishable states equals:

$$W = \frac{N!}{n_1!, n_2! \cdots}$$

From this, if the frequencies n_i are large, so that $n_i! \approx n_i^{n_i}$ holds, for the Boltzmann entropy (with relative frequencies $p_i := n_i/N$) the following it true:

$$\begin{aligned}\log(W) &= \log\left(\frac{N!}{n_1!, n_2! \cdots}\right) \approx N\log(N) - \sum_i n_i \log(n_i) = N\sum_i \frac{n_i}{N}\log(N) - N\sum_i \frac{n_i}{N}\log(n_i) \\ &= -N\sum_i \left(\frac{n_i}{N}\right)\log\left(\frac{n_i}{N}\right) = -N\sum_i p_i \log(p_i)\end{aligned}$$

In summary:

- The derivation follows from the classification of the particles implied by expansion in space as well as from the Boltzmann formula and application of the counting principle to the *total* number of states.

---

[8] Cf. section 3.5.2 on the issues of the counting principle

[9] Cf. section 4.2 on individuals and collectives

[10] https://en.wikipedia.org/wiki/Entropy_of_mixing#Proof_from_statistical_mechanics

- The derivation does not follow from a classification of the states and counting per class. Accordingly, the relative frequencies p_i are not those of the states, but particle frequencies.

According to its construction, the described system can be decomposed into subsystems, the cells C_i. Now the determination of the number of states within the cells is subject to the same condition as that for the total system. And this condition is the distinguishability of the particles, from which follows that each cell has the number of states 1, since there are also no distinguishable states in it because of the indistinguishability of the particles. From this again follows from the product rule of the counting principle that there is also in the whole system only one state and the entropy of the whole system like that of each cell is equal to 0.

Against this new result, which contradicts the old $-Np_i \log(p_i)$, one may object that the independence of the states of different cells cannot be clarified and thus the numbers of states cannot simply be multiplied. However, this objection can be excluded, because possible dependencies at best reduce the total number of states, but this again is not possible, because the product rule comes to only one state of the total system anyway.

Another objection may be, according to which there is indeed only one total state, because the interchange of particles e.g. by transposition does not change it. This is true for interchanges, but not for particle displacements, by which e.g. from a uniform distribution distinctive uneven distributions can arise, with which one gets at the same time into the contrast between the expansion and the contraction of matter in space as well as in momentum space, which is one of those laws, that define the meaning of entropy, i.e. the density entropy and the entropy of motion respectively.

In a further summary we can state firstly that with the counting principle and counting of all states from the Boltzmann entropy one can derive an expression which formally agrees with the Gibbs entropy – except for the factor of the total number of particles. Second, it suggests that the product rule of the counting principle for determining entropy from the entropies of subsystems does not have general validity in the composition of parts into a whole, precisely in that context in which it lends itself as an explanatory ground for the extensivity of entropy. As will be shown, these problems are related to the fact that entropy essentially also reflects the globalization effort[11] which arises from the composition of parts to a whole and which can only be assigned to the whole and not to the parts, while the counting principle, on the other hand, assigns to each of the parts a quantity – the number of states – which all together supposedly explain the entropy of the whole. Instead, it seems that the whole is more than the sum of its parts. In connection with this, I would like to put forward the following theses:

Boltzmann Entropie: The Boltzmann entropy 1.2 is incomplete. If one derives from it the entropy $-N \sum p_i \log(p_i)$ including the total particle number N as above, it still lacks – in respect to the property of extensivity – the complexity of the continuum.[12]

Gibbs Entropie und Zähl-Prinzip: Gibbs entropy is not extensive because the product rule of the counting principle has only limited and, as will be shown, only fuzzy validity, while at the same time the extensivity of Gibbs entropy can only be explained by the product rule.[13]

Especially in section 7.6 Gibbs entropy (Shannon entropy) is compared with the later defined density entropy or the U-entropy, which in contrast to the normal understanding of entropy is a measure of disequilibrium and contraction, and it becomes clear that the finiteness of entropy is related to the double reference of two continuous sets, the particle continuum on the one hand and the space continuum on the other hand. It is equally clear that, in contrast, Gibbs entropy

---

[11] Cf. section 3.3.2

[12] Cf. section 3.4 on the complexity of spacial continuum

[13] Cf. section 3.5.2 on the issues of extensivity of Shannon entropy

is self-referential and just the complexity of only one continuous set. The counting principle is obviously subject to self-reference in the same way, because it basically counts the states of only one set, namely those of the state space.

Newton's *actio* = *reactio* is perhaps the most famous example of a double reference in physics, where two bodies reference each other. Another example is the twin currents of the Schrödinger equation, while the heat conduction equation and the diffusion equation of the isolated system are both self-referential. Besides the already above mentioned double reference between particles and physical space, then possibly another double reference is that between kinematics and dynamics in such a way that the dynamics for each particle determines the evolution from which its probability distribution emerges and thus also the collective particle density, which as a second continuum is the one basis of the mutual reference between space and matter. This duality in turn forms the basis for the finiteness of entropy, and it is at the same time the basis of the hierarchical, i.e. the intensive order of space, which is reflected in the definition first of the rational paths, which arise through the hierarchy as distinguished paths and then points of the continuous space and thus, together with the irrational paths, make the space an address space, which conversely now forms the basis for the dynamics of the particles in the form of the Schrödinger equation, in that this equation refers to the addresses defined by the kinematics. With the kinematics it is about nothing else than the bare description of state, while the dynamics define the how and why of the change of state. Entropy belongs to kinematics.

Perhaps the most important implication of the considerations presented here is the discontinuity of the density function of the particle density, which is even not only discontinuous, but probably nowhere defined – with infinite values of the density for the rational and bounded but divergent densities for the irrational paths, which together justifies the title of the book – the infinite noise. On the other hand, it is integrable everywhere, since the mean values exist at all scales. Thus, the mean is also another defining concept of the book, which is directly related to the two concepts of expansion and contraction already mentioned. For any deviation of a distribution from the distribution mean implies at the same time the existence of a contraction, while the internal expansion in a system – of restricted extensive size – is complete when the distribution coincides with the uniform distribution. In extensively unrestricted systems but with given finite standard deviation, the – then exclusively – internal expansion reaches a maximum if the system is normally distributed.

This is not the only reason why extensivity and intensivity form another outstanding pair of concepts. Entropy is expected to be extensive. It is shown here that it is a physical quantity with extensive and at the same time intensive characteristics, in that it reflects not only the external but also the internal quantity of a system. In this sense, ordinary entropy is also an expansion measure, increasing with both external and internal expansion. In contrast, the U-entropy defined in chapter 7 is a contraction and order measure, which is the larger, the more exposed parts of the system appear. In this way the order of the cosmic universe originates from the galaxies and fixed stars which are excellent and exposed places of space.

Finally, however, the book is not exclusively about these concepts, for the understanding of which only chapters 3, 6 and 7 are necessary, while the remaining chapters describe side paths in the context of entropy.

Chapter 2

The historical and further physical context

2.1 Expansion and contraction in the Carnot cycle

The *Carnot cycle*[1] was apparently the first major milestone on the way to the formulation of the Second Law. In this thought experiment by *Nicolas Léonard Sadi Carnot*, a gas is transformed from an initial state into other states one after the other in a process consisting of four sub-processes, but to return to the initial state at the end, which is why the whole process is also called a circular process. The focus of the process is the expansion and contraction of a gas, especially the isothermal sub-processes, i.e. the expansion and contraction of the hot and cold gas, respectively, at constant temperature, which is only possible under heat exchange with the environment, a heat pool. In comparison, the other two are only auxiliary processes, again one an expansion and the other a contraction process, but this time under isolation conditions, i.e. excluding heat exchange with an environment. These two sub-processes are needed so that the return to the initial state is possible and the whole process is actually a circular process, which in turn allows the establishment of an energy balance with respect to the two isothermal sub-processes, which is the actual goal of the circular process. More precisely, the four sub-processes are described, for example, in *Carnot cycle* as follows:

sothermal expansion (with heat exchange): the initially hot gas with temperature T_H and volume V_1 at pressure P_1 is expanded to the larger volume V_2 with lower pressure P_2 by using a heat pool, the environment of the gas, while the temperature is kept constant.

Isentropic expansion (without heat exchange): The gas is adiabatically reversibly expanded, i.e. without using a heat sink and thus without heat exchange with an environment, further to the volume V_3 with pressure P_3, cooling down to temperature T_L.

Isothermal compression (with heat exchange): The gas is isothermally compressed to the volume V_4 at pressure P_4 using a heat sink of temperature T_2.

Das Gas wird unter Verwendung eines Wärmebeckens der Temperatur T_2 isotherm auf das Volumen V_4 bei einem Druck P_4 komprimiert.

Isentropic compression (without heat exchange): The gas is adiabatically reversibly compressed further and thus heated until it returns to its initial state with state values V_1, P_1 and temperature T_H.

---

[1] https://en.wikipedia.org/wiki/Carnot_cycle

As already mentioned, the adiabatic sub-processes have a bridging function by bringing the gas alternately to the two working temperatures T_H and T_L of the isothermal processes, within which the work done mechanically on the gas $P(+dV)$ or by the Gas $P(-dV)$ also only has a bridging function, since it is assumed that it only mediates the heat exchange between the gas and the heat pool. I.e. it does not actually occur itself. Instead heat exchange takes place without time delay. After all, the thus fictitiously assumed mechanical work performed by the gas or the heat pool can be calculated and thus ultimately enables the energy balance mentioned above to be drawn up.

The two *adiabatic processes*[2] satisfy the following condition:

$$TV^{\gamma-1} = const. \tag{2.1}$$

So also $T_H V_2^{(\gamma-1)} = T_L V_3^{(\gamma-1)}$ and $T_L V_4^{(\gamma-1)} = T_H V_1^{(\gamma-1)}$ and thus $V_2/V_3 = V_1/V_4$ bzw. $V_2/V_1 = V_3/V_4$ and therefore:

$$\log\left(\frac{V_2}{V_1}\right) = -\log\left(\frac{V_4}{V_3}\right)$$

The last equation, in turn, relates the two isothermal processes, each of which involves notional work output and transfer of energy to the other system. In the case of isothermal expansion, it must be the gas that absorbs energy, while the heat pool is the system doing the work. And from the fiction of mechanical transfer, it follows that $\int_{V_1}^{V_2} p\,dV$ is the quantity of work done that is directly absorbed by the gas in the form of heat. Thus, if Q_H is this quantity of heat, the following holds true:

$$Q_H = \int_{V_1}^{V_2} p\,dV$$

In isothermal compression, the gas loses energy analogously:

$$Q_L = \int_{V_3}^{V_4} p\,dV$$

In summary, for each of the two isothermal processes, the gas and the heat pool exchange heat even though they have the same temperature. The exchange is mediated by the mechanical work done in each case, which one system does and loses heat energy in the process, and which the other system absorbs and converts directly into heat energy.

Finally, the *ideal gas law* applies to the gas:[3]

$$pV = nRT = nk_B N_A T \tag{2.2}$$

Since in both isothermal processes the temperature remains constant, taking into account the above equations, we get:

---

[2] https://en.wikipedia.org/wiki/Adiabatic_process

[3] https://en.wikipedia.org/wiki/Ideal_gas_law

$$\frac{Q_H}{Q_L} = \frac{\int_{V_1}^{V_2} p\,dV}{\int_{V_3}^{V_4} p\,dV} = \frac{nRT_H \int_{V_1}^{V_2} \frac{1}{V}\,dV}{nRT_L \int_{V_3}^{V_4} \frac{1}{V}\,dV} = \frac{T_H}{T_L} \frac{\log\left(\frac{V_2}{V_1}\right)}{\log\left(\frac{V_4}{V_3}\right)} = -\frac{T_H}{T_L}$$

All in all, this gives the condition for the four-part process to be closed, i.e. to return to the initial state:

$$\frac{Q_H}{T_H} + \frac{Q_L}{T_L} = 0 \tag{2.3}$$

Now the expressions Q/T form the basis for the entropy definition according to *Rudolf Clausius*. The absorption of the heat quantity Q_H from the environment at temperature T_H is connected with the entropy increase Q_H/T_H, correspondingly the removal of Q_L at T_L with the entropy decrease Q_L/T_L. In each case, the reverse is true for the environment, so that the complete entropy balance is zero for both the gas and the environment. This is possible by the fiction of reversible heat transfer between gas and environment avoiding a temperature difference between both systems. Exactly this condition, heat exchange between two systems without a temperature gradient between them, is not fulfilled in reality. On the contrary, here the heat flows from the warm to the cold system. Not the sum of entropy increase of one system and entropy decrease of the other system is equal to zero, but in general the sum of inflow and outflow of heat. Instead of Eqn. 2.3, the entropy change ΔS of the total system consisting of both subsystems equals, if $T_H > T_L$ is valid and the heat flux δQ is small:

$$dS = \delta Q \left(\frac{1}{T_L} - \frac{1}{T_H}\right) \tag{2.4}$$

The cold system records the heat input $\delta Q > 0$ at temperature T_L and thus the entropy increase $\delta Q/T_L$, correspondingly the warm system records the entropy decrease $-\delta Q/T_H$. The sum of the entropy changes for the total system is greater than zero as long as the two temperatures are different. However, equation 2.4 arbitrarily sets the sign for δQ from the perspective of the cold system. A perspective-free symmetric formulation for two systems with arbitrary temperatures T_1 and T_2 and energy fluxes δQ_1 and δQ_2 is

$$dS = \left(\frac{\delta Q_1}{T_1} + \frac{\delta Q_2}{T_2}\right) \geq 0 \tag{2.5}$$

with the guarantee for the non-negativity of the sum by the auxiliary conditions of the two theorems of thermodynamics, the *first law of thermodynamics*[4], i.e. the conservation of energy

$$\delta Q_1 + \delta Q_2 = 0 \tag{2.6}$$

and the second theorem:

$$\operatorname{sgn}(\delta Q_1) = \operatorname{sgn}(T_2 - T_1) \tag{2.7}$$

---

[4] https://en.wikipedia.org/wiki/First_law_of_thermodynamics

2.2 The Second Theorem

2.2.1 Infinitesimal locality of two-part systems

Exchange and net entropy

If one imagines a system, e.g. a solid, composed of many very small parts, then the definition for the (heat related) entropy change of a closed system in the usual notation is

$$\Delta S = \int \frac{\delta Q}{T} \geq 0 \tag{2.8}$$

where first over the subsystems and then the time has to be integrated.[5]

At first sight, Eqn. 2.8 for the many-part system seems to be a direct generalization of Eqn. 2.5 for the two-part system, for which the isolation is implicitly assumed in Eqn. 2.5, because only the two subsystems are considered. The general statement of monotonicity in Eqn. 2.8 for the many-part system is mathematically not obvious and therefore only a postulate, while the implicit isolation of the two-part system in connection with the two auxiliary conditions of the theorems 2.6 and 2.7 enforces the inequality in 2.5 mathematically. More precisely, the following holds for the (isolated) two-part system:

Under the condition of conservation of energy 2.6 ($\delta Q_1 + \delta Q_2 = 0$) the monotonicity of entropy in 2.5 ($dS = \delta Q_1/T_1 + \delta Q_2/T_2 \geq 0$) is equivalent to the *second law* according to 2.7 ($\mathrm{sgn}(\delta Q_1) = \mathrm{sgn}(T_2 - T_1)$), because the equivalence

$$dS = \delta Q_1 \left(\frac{1}{T_1} - \frac{1}{T_2} \right) = \delta Q_1 \left(\frac{T_2 - T_1}{T_1 T_2} \right) \geq 0 \Leftrightarrow \mathrm{sgn}(\delta Q_1) = \mathrm{sgn}(T_2 - T_1)$$

holds. I.e. also the two statements

1. Heat always flows from the hot to the cold system (Eqn. 2.7)
2. The entropy of a closed system does not decrease (Eqn. 2.5)

are equivalent for the two-part system. Now, both statements are different formulations of the *second law*[6], whose connection is mathematically revealed in a very simple way in the two-part system. However, one can put forward the thesis that the connection for the two subsystems and their compound is simple only under one essential condition, which consists in the fact that in each of them the temperature is distributed uniformly in space. For temperature averages T_1 and T_2, for example, Eqs. 2.5 und 2.7 do not hold in general, or the statements are true, for example, only under the assumption of an indefinite time horizon. For arbitrary and thus also arbitrarily short time horizons, however, spatially uniformly distributed temperatures are required, which again can only be assumed in general for spatially arbitrarily small systems. In this respect, the second law in the formulation (1) is afflicted with fuzziness, which can only be avoided by arbitrarily large periods of time or arbitrarily small systems.

The formulation (2), however, is assumed to hold, first, with certainty, second, for all – especially short time – periods, and third, for spatially extended systems, and finally, the magnitude of its increase is subject to the inequality 2.8.

According to what has been said, (1) is the differential formulation of the second law in the limit of arbitrarily small systems and (2) is the integral for extended systems. In that case the integral

---

[5] Cf. https://en.wikipedia.org/wiki/Disgregation

[6] https://en.wikipedia.org/wiki/Second_law_of_thermodynamics

in Eqn. 2.8 must be equal to the integral over the differential relations 2.7, which will be shown in the following:

For this purpose, the two-part system with equations 2.5, 2.6 and 2.7 is considered as the basis of the differential relationship between two neighboring subsystems of the many-part system. The physical assumption here is that the interaction of the two subsystems – resulting in heat exchange – takes place at their interface (or at least through their interface or defined with respect to it). Eqn. 2.5 is thus given a second interpretation. In the previous interpretation, the heat flux δQ_1 in 2.6 has a reference to the first subsystem and only to it, and correspondingly δQ_2 to the second subsystem, just as δQ in 2.8 has a reference to exactly one subsystem with the meaning of the net heat flux into or out of that subsystem resulting from the interactions of the subsystem with all its neighbors, so that δQ in 2.8 is the sum e.g. of six such boundary heat fluxes if one visualizes, say, the subsystems as cubes, each of which is then separated from its neighbors by exactly six interfaces. In the second interpretation, there is only one boundary heat flux (hereafter δq) for the two-part system, which is to be distinguished from the net heat fluxes (hereafter δQ) and has a reference only to the two-part system. Instead of the equations 2.5, 2.6 and 2.7, then the following holds:

$$dS = \delta q \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \geq 0 \tag{2.9}$$

$$\operatorname{sgn}(\delta q) = \operatorname{sgn}(T_2 - T_1) \tag{2.10}$$

With this interpretation, the distinction of the two subsystems fades away, since the equations together are invariant to the interchange of the subsystems. This becomes clearer, if from 2.9 and 2.10

$$dS = |\delta q| \, \frac{|T_2 - T_1|}{T_1 T_2} \geq 0 \tag{2.11}$$

is derived, which expresses the complete symmetry of the two subsystems with respect to their total entropy increase. This in turn has the further consequence that the entropy increase of the two-part system is independent of its relation to the many-part system in which it is embedded, whereby the entropy within the many-part system is additive. If, however, as a third consequence, the entropy increase of a whole system is the cumulative result of the inner relations of its two-part subsystems, then there is an entropy in principle only for extensively unbounded systems, since the two-part relation founded by neighborhood – analogous to the neighborhood or successor relation $n \to n+1$ of the natural numbers – implies unbounded expansion. Thus, at this point, one can already recognize the meaning of the isolation of the physical system for the entropy, since the isolation, by preventing the heat exchange on a finite system periphery, sets a limit to the expansion of the physical connection, which is in itself unlimited, and thus makes the summation of the two-part entropy increases 2.11 on a limited (closed) area meaningful at all.

The thus monotonic entropy change dS of the two-part system according to 2.9 or 2.11 shall be called *exchange entropy* in the following while the entropy change corresponding to the net heat flux of each of the subsystems of the two-part system shall be called *net entropy*:

Exchange entropy: The non-negative exchange entropy is the entropy change according to equation 2.11 of a total system consisting of two parts. It is by definition non-negative in accordance with the *first law* and the *second law* according to (1). I.e., the exchange entropy is an expression of the universal validity of the two laws. It is additive with the consequence of extensive unboundedness of the physical system or of an induced boundedness, which we call isolation.

Net entropy: The net entropy (change) is the entropy change of a subsystem due to the interaction with all its neighboring subsystems. It is the sum of the exchange entropies, each resulting from the boundary heat fluxes between the subsystem and its neighbors. Unlike the exchange entropy, it can also be negative. It can be directly generalized to many-part systems, as done in Eqn. 2.8, i.e. the entropy change of a many-part system is understood as the sum of the net entropies of its subsystems.

Further, we shall now show that for isolated systems the sum of the exchange entropies coincides with the sum of the net entropies, so that since the exchange entropies are non-negative without exception, the total entropy $\int \delta Q/T$ cannot decrease either.

Decisive for this global relation between the exchange entropies and the net entropies is the identity of the sums of two series which are rearrangements from one another. Let the numbers $a_0, a_1, \ldots a_n$ und $b_0, b_1, \ldots b_n, b_{n+1}$ with the boundary condition

$$a_0 b_0 = a_n b_{n+1} \tag{2.12}$$

be given. Because of

$$\sum_{i=0}^{n} a_i(b_i - b_{i+1}) = \sum_{i=0}^{n} a_i b_i - \sum_{i=1}^{n+1} a_{i-1} b_i = a_0 b_0 + \sum_{i=1}^{n} (a_i - a_{i-1}) b_i - a_n b_{n+1} \tag{2.13}$$

then applies:

$$\sum_{i=0}^{n} a_i(b_i - b_{i+1}) = \sum_{i=1}^{n} (a_i - a_{i-1}) b_i \tag{2.14}$$

Thus, the generally valid equality in 2.14 is equivalent to the equality in 2.12.

Now let further be given a rod which is decomposed into n cells $C_1, \ldots C_n$ with temperatures $T_1, \ldots T_n$, internal boundary heat fluxes δq_i between two neighboring cells C_i and C_{i+1} for $1 \leq i \leq n-1$ and the net heat fluxes:

$$\delta Q_i = \delta q_i - \delta q_{i-1} \tag{2.15}$$

Let the boundary heat fluxes at the rod ends C_1 und C_n with the surroundings of the rod be δq_0 and δq_n, respectively. All together the following table shows

C_0	C_1	C_2	...	C_n	C_{n+1}
T_0	T_1	T_2	...	T_n	T_{n+1}
$\delta q_0 \rightarrow$	$\delta q_1 \rightarrow$	$\delta q_2 \rightarrow$	...	$\delta q_n \rightarrow$	
	$\rightarrow \delta Q_1 \leftarrow$	$\rightarrow \delta Q_2 \leftarrow$	...	$\rightarrow \delta Q_n \leftarrow$	

Table 2.1: Heat flow in the rod

where the additional cells C_0 and C_{n+1} represent the environment at the front and rear ends of the rod, respectively. For these two cells, the net heat fluxes δQ_0 and δQ_{n+1}, respectively, are missing from the figure because we are interested in the net heat fluxes only within the rod. Similarly, the boundary heat flux δq_{n+1} is missing because it is not directly related to the rod. From Eqn. 2.14 with constraint 2.12, it follows with $a_i = \delta q_i$ und $b_i = 1/T_i$

$$\sum_{i=1}^{n} \delta q_i \left(\frac{1}{T_i} - \frac{1}{T_{i+1}} \right) = \sum_{i=1}^{n} (\delta q_i - \delta q_{i-1}) \frac{1}{T_i} \tag{2.16}$$

together with the transferred boundary condition:

$$\delta q_0 T_0 = \delta q_n T_{n+1} \tag{2.17}$$

The symmetry and positivity of the exchange entropy, expressed primarily in Eqn. 2.11, allows the boundary heat fluxes δq_i to be given an arbitrary (but consistent) direction (indicated in Table 2.1 above by the direction of the arrows in line with the progression of the index i and thus consistent with the traversal sense of the spatial order). From Eqn. 2.9 follows:

$$\operatorname{sgn}(\delta q_i) = \operatorname{sgn}(T_{i+1} - T_i) \tag{2.18}$$

The satisfaction of constraint2.17 can now be related to the isolation of the system as follows:

Open rod: rod with open ends for which there is no heat exchange at either end with its surroundings, and for which therefore $\delta q_0 = \delta q_n = 0$ is true.

Closed rod: Closed rod, e.g. a ring. For this, the three cells C_0, C_n and C_{n+1} are identical, so that $T_0 = T_n = T_{n+1}$ and $\delta q_0 = \delta q_n$, and hence 2.17 also holds.

Thus, in both cases, equation 2.16 is satisfied and, by premise, its left-hand side consists of the non-negative exchange entropies only, so that from 2.16 together with Eqn. 2.8 and 2.15

$$0 \leq \sum_{i=1}^{n} \delta q_i \left(\frac{1}{T_i} - \frac{1}{T_{i+1}} \right) = \sum_{i=1}^{n} (\delta q_i - \delta q_{i-1}) \frac{1}{T_i} = \sum_{i=1}^{n} \frac{\delta Q_i}{T_i} = \Delta S \tag{2.19}$$

follows. Furthermore, due to

$$\sum_{i=1}^{n} \delta Q_i = \sum_{i=1}^{n} (\delta q_i - \delta q_{i-1}) = \delta q_n - \delta q_0 = 0 \tag{2.20}$$

the *first law* ist satisfied, if also the boundary condition $\delta q_0 = \delta q_n$ is satisfied.

Equation 2.15 is nothing else than the *continuity equation*[7] for the heat energy, i.e. the local version of the law of conservation of energy. As one can see clearly, the conservation of energy of the whole system does not follow in general, but – here – only under the additional condition $\delta q_0 = \delta q_n$. This is clear, because the global conservation of energy can follow from the local conservation only if the local conservation has global validity in the system. For systems, whose spatial extension is limited, and which have therefore a periphery, this plays therefore the crucial role for the global conservation of energy. In this context, the isolation of the system means that the local conservation of energy also applies to the peripheral cells.

---

[7] https://en.wikipedia.org/wiki/Continuity_equation#Energy_and_heat

Fourier's law and its consequences

One can now go one step further and develop the second law in the form of Equations 2.9 und 2.10 with the help of *Fourier's law*[8], since the law establishes a relationship between the boundary heat flux and the temperature or the temperature gradient. It should be taken into account that Fourier's law describes a heat flux in the direction of decreasing temperature, but the sign of δq_i is defined inversely according to Eqn. 2.18, namely the positive heat inflow into cell i when cell $i+1$ has a higher temperature, so that in the limit $dx \to 0$ with the thermal conductivity k

$$\delta q_i = - - k \Delta y \Delta z \frac{\Delta T}{\Delta x} \Delta t = k \frac{T_{i+1} - T_i}{\Delta x} \Delta y \Delta z \Delta t \longrightarrow k \frac{\partial T}{\partial x} \Delta y \Delta z \Delta t \tag{2.21}$$

applies. It follows from this:

Size of the exchange entropy: Because in the limit $dx \to 0$ also

$$\frac{1}{T_i} - \frac{1}{T_{i+1}} = \frac{T_{i+1} - T_i}{T_i T_{i+1}} = \frac{\Delta x}{T_i T_{i+1}} \frac{\Delta T}{\Delta x} \longrightarrow \frac{1}{T^2} \frac{\partial T}{\partial x} dx$$

holds, together with Eqn. 2.21 for the exchange entropies in Eqn. 2.19 is obtained:

$$\delta q_i \left(\frac{1}{T_i} - \frac{1}{T_{i+1}} \right) \longrightarrow k \frac{\left(\dfrac{\partial T}{\partial x} \right)^2}{T^2} dx \Delta y \Delta z \Delta t$$

Now this is the exchange entropy of cell i due to the rearward boundary heat flux between it and cell $i+1$. In an extension to three dimensions, the rearward exchange entropies in y-direction and in z-direction are to be considered analogously. Namely, if the system is decomposed into cubes of equal size along the coordinate axes within the framework of a Cartesian coordinate system, then this decomposition also specifies decompositions into square rods in each of the x-direction, y-direction and z-direction. For each such bar, relations as in 2.19 then hold, and each cube ijk is the intersection of three bars. For the sum

$$ex_{ijk} := \delta q_{ijk,x} \left(\frac{1}{T_{ijk}} - \frac{1}{T_{(i+1)jk}} \right) + \delta q_{ijk,y} \left(\frac{1}{T_{ijk}} - \frac{1}{T_{i(j+1)k}} \right) + \delta q_{ijk,z} \left(\frac{1}{T_{ijk}} - \frac{1}{T_{ij(k+1)}} \right)$$

of the three rearward exchange entropies of the cube holds if the side lengths of the cube converge to zero, with $dV := dx \cdot dy \cdot dz$:

$$ex_{ijk} \longrightarrow ex(\mathbf{r})\, dV = k \frac{1}{T^2} \left(\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right) dV \Delta t$$

or short:

$$ex(\mathbf{r})\, dV = k \left(\frac{\nabla T}{T} \right)^2 dV \Delta t \tag{2.22}$$

---

[8] https://en.wikipedia.org/wiki/Thermal_conduction#Fourier's_law

Magnitude of net heat flux: For the net heat flux $\delta Q_i = \delta q_i - \delta q_{i-1}$ of a bar according to 2.15, we obtain from Eqn. 2.21:

$$\delta Q_i \longrightarrow k \frac{\frac{\partial T(x)}{\partial x} - \frac{\partial T(x-dx)}{\partial x}}{dx} dx\Delta y\Delta z\Delta t = k\frac{\partial^2 T}{\partial x^2} dx\Delta y\Delta z\Delta t$$

For the three-dimensional extension

$$\delta Q_{ijk} := (\delta q_{ijk,x} - \delta q_{(i-1)jk,x}) + (\delta q_{ijk,y} - \delta q_{i(j-1)k,y}) + (\delta q_{ijk,z} - \delta q_{ij(k-1),z}) \tag{2.23}$$

follows accordingly:

$$\delta Q(\mathbf{r})dV = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) dV\Delta t = k \cdot \Delta T \cdot dV \cdot \Delta t \tag{2.24}$$

Size of the net entropy: For the net entropy according to Eqn. 2.19, i.e.

$$net_{ijk} := \frac{\delta Q_{ijk}}{T_{ijk}}$$

follows from 2.24 in the limit of infinitesimal cell size $\Delta x\Delta y\Delta z \longrightarrow dV$:

$$net(\mathbf{r})\, dV = k\frac{\Delta T}{T} \cdot dV \cdot \Delta t \tag{2.25}$$

Isolated systems

The results in Equations 2.22 and 2.25 for the exchange and net entropies, respectively, follow solely from *Fourier's law* in Eqn. 2.21, which concretizes the second law by relating the boundary heat flux δq_i to the temperature gradient $\partial T/\partial x$.

The energy conservation law 2.20 of the isolated one-dimensional system can be extended to three-dimensional systems in a simple way by identifying isolation with extensive finiteness, so that there is no heat exchange with the environment on the whole system periphery, which is also the case if the perpendicular component of the temperature gradient T on the periphery disappears everywhere. For this case Eqn. 2.16 is valid and correspondingly also the three-dimensional expansion

$$0 \leq \sum_{ijk} ex_{ijk} = \sum_{ijk} net_{ijk} = \Delta S$$

and therefore according to equations 2.22 and 2.25 also

$$0 \leq \Delta t \int k\left(\frac{\nabla T}{T}\right)^2 dV = \Delta t \int k\frac{\Delta T}{T} dV = \int \frac{\delta Q}{T} = \Delta S$$

and finally:

$$\frac{dS}{dt} = \int k\frac{\Delta T}{T} dV = \int k\left(\frac{\nabla T}{T}\right)^2 dV \geq 0 \tag{2.26}$$

For the open bar[9] with length L, if k is independent of location, the equation from 2.26 becomes:

$$\frac{dS}{dt} = \int_0^L \frac{T''}{T}\,dx = k\int_0^L \left(\frac{T'}{T}\right)^2 dx \geq 0 \tag{2.27}$$

With the boundary condition $\delta q_0 = \delta q_n = 0$ applying to the open rod, from which follows $T'(0) = T'(L) = 0$, the equality of the two integrals above also follows directly from the product rule:

$$\int_0^L \frac{T''}{T}\,dx = \frac{T'(L)}{T(L)} - \frac{T'(0)}{T(0)} - \int_0^L T'\left(-\frac{1}{T^2}T'\right) dx = \frac{T'(L)}{T(L)} - \frac{T'(0)}{T(0)} + \int_0^L \left(\frac{T'}{T}\right)^2 dx \tag{2.28}$$

For the closed rod[10] Eqn. 2.27 also holds, but with boundary conditions $T(0) = T(L)$ und $T'(0) = T'(L)$, from which the equality of the integrals in 2.27, as for the open rod, also follows directly from the product rule, as is clear from Eqn. 2.28.

For the bar, one can also derive the time monotonicity of the entropy more generally than by its complete isolation from Eqn. 2.28 if

$$\frac{T'(L)}{T(L)} - \frac{T'(0)}{T(0)} \geq 0$$

holds, because then instead of 2.27 even

$$\frac{dS}{dt} = \int_0^L \frac{T''}{T}\,dx \geq k\int_0^L \left(\frac{T'}{T}\right)^2 dx \geq 0 \tag{2.29}$$

follows. A simple concrete case is the heat flow with heat inflow at the back end and heat outflow of the same magnitude at the front end, for which $T'(0) = T'(L) > 0$ thus holds, but which encounters an opposite global temperature drop $T(0) \geq T(L)$.

Now it is interesting in all three cases that because of the boundary conditions the integral over the relative temperature curvature T''/T is guaranteed to be non-negative, although the curvature can also be negative regionally, since the boundary conditions do not directly predetermine the temperature profile in the interior of the interval. However, the boundary condition $T'(0) = T'(L)$, which is present in all three cases, first forces the integral over the absolute curvature T''

$$\int_0^L T''\,dx = T'(L) - T'(0) = 0$$

and thus the mean value over all curvatures vanishes. It must therefore be assumed that the relative curvature magnitudes $|T''/T|$ tend to be small in right-curved (concave) regions with $T'' \leq 0$ and thus the temperatures tend to be high and, conversely, large in left-curved (convex) regions with correspondingly low temperatures. This is obviously true for well-known functions such as the sine between $-\pi/2$ and $+\pi/2$ or between $\pi/2$ and between $3\pi/2$. In general, maxima

---

[9] Cf. the definition of the open rod in section 2.2.1

[10] Cf. the definition of the closed rod in section 2.2.1

are indeed surrounded by right-hand curvatures and minima by left-hand curvatures, confirming the conjecture expressed above. In summary, the boundary conditions and thus the isolation of the system therefore do not determine the course of the temperature inside the system, but they do have a decisive and, above all, global influence. If a system is isolated ad hoc, the temperature distribution must therefore globally satisfy the conditions of entropy increase within a short time, which in one extreme case requires fast adjustments of global extent and in the other extreme case means that the system or large parts of the system were already practically isolated before.

Further substituting on the left-hand side of Eqn. 2.26 ΔT using the heat conduction equation 2.38, with a reference temperature T_R and the integration constant $-\log(T_R)$, we obtain

$$\frac{dS}{dt} = \int k \frac{\Delta T}{T} \, dV = \int \frac{k}{\alpha} \frac{\frac{\partial T}{\partial t}}{T} \, dV = \int \rho c \frac{\partial}{\partial t} \left(\log \left(\frac{T}{T_R} \right) \right) dV \geq 0$$

and thus finally for temporally and spatially constant density ρ and heat capacity c:

$$\frac{dS}{dt} = \rho c \frac{d}{dt} \int \log \left(\frac{T}{T_R} \right) dV \geq 0 \tag{2.30}$$

The monotonicity($\geq$) in 2.30 also assumes the validity of the right-hand part of Eqn. 2.26, i.e., also the global conservation of energy. Altogether we have assumed:

- constant volume
- constant energy
- density and heat capacity constant in time and space

Under these conditions, we can finally derive from Eqn. 2.30:

$$S = \rho c \int \log \left(\frac{T}{T_R} \right) dV \tag{2.31}$$

Due to Eqn. 2.39, the conservation of energy is equivalent to

$$0 = \frac{dE}{dt} = \frac{d}{dt} \int \delta Q \, dV = \int \frac{\delta Q}{\partial t} \, dV = c\rho \int \frac{\partial T}{\partial t} \, dV = c\rho \frac{d}{dt} \int T \, dV$$

and thus equivalent to the constancy of the temperature mean T_M, namely:

$$\frac{dT_M}{dt} = \frac{d}{dt} \frac{1}{V} \int T \, dV = 0 \tag{2.32}$$

Further, for the integral in Eqn. 2.31, it follows from Jensen's inequality 3.30:

$$\int \log \left(\frac{T}{T_R} \right) dV = V \int \log \left(\frac{T}{T_R} \right) \frac{dV}{V} \leq V \log \left(\frac{1}{V} \int \frac{T}{T_R} \, dV \right) = V \log \left(\frac{T_M}{T_R} \right)$$

So it is also:

$$S \leq S_{max} := \rho c V \log \left(\frac{T_M}{T_R} \right) \tag{2.33}$$

In the last two inequalities, equality holds exactly when all temperatures are equal, i.e., in the case of uniform distribution for which all temperatures in the system coincide with the temperature mean T_M. Since S_{max} is time independent according to Eqn. 2.32, it follows that entropy stops increasing in any case when the temperature in the system is uniformly distributed, which we identify with thermodynamic equilibrium.

So, in a summary, we can say that when the energy of the system is constant, the following holds:

- The volume integral over temperature and hence the temperature average are also constant in time (Eqn. 2.32).
- The volume integral over the logarithm of the temperature is not necessarily constant in time (Eqn. 2.30), but it is when the temperature is uniformly distributed, which is the thermodynamic equilibrium (Eqn. 2.33 and 2.32), at which the entropy reaches its maximum value.

Relations to the Third Theorem

Finally, it is noticeable that the results 2.26, 2.30 and 2.33 seem to have no validity at the absolute temperature zero $T = 0$, which would be in clear contradiction to the *third law*[11], according to which the entropy of a system in thermodynamic equilibrium has a definite and in any case finite value. For equations 2.30 and 2.33 this is clear because the logarithm does not exist at $T = 0$. But the same is true for the right side of Eqn. 2.26 because of $T^{'}/T = \log(T)^{'}$.

The resolution of the contradiction lies, as described in *Third law of thermodynamicd - Specific heat*[12], in the temperature dependence of the specific heat capacity, for Fermi gases and for Bose gases, respectively:

$$C_V \sim T$$

$$C_V \sim T^{3/2}$$

The temperature dependence of the specific heat in thermodynamic equilibrium is most easily and clearly noticeable according to eqn. 2.33. In the case of Fermi particles, e.g.

$$S_{max} \sim T_M \log\left(\frac{T_M}{T_R}\right)$$

which also converges to 0 in the limit $T_M \to 0$.

2.2.2 Globality and finite two-part systems

The core of the argumentation of the last section can be countered by a lack of reality. The core is that the two-part system is irrelevant for entropy as long as it has finite size. Now, instead, one can also argue the antithesis, according to which every two-part system and the global relations between the two parts of even arbitrary finite size are at least related to the actual entropy of the system. In this idea, the two temperatures T_1 and T_2 represent mean values of the respective subsystem, on the basis of which equations 2.5, 2.6 and 2.7 or, say, Eqn. 2.11 provide an initial estimate of the actual entropy increase of the overall system. This is exactly what we all do when we find gross imbalances for a system, e.g., that between a low pressure and a high pressure area,

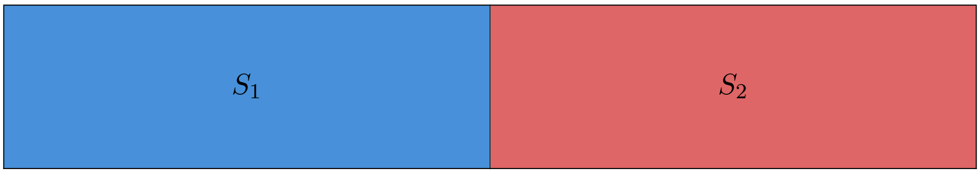
---

[11] https://en.wikipedia.org/wiki/Third_law_of_thermodynamics

[12] https://en.wikipedia.org/wiki/Third_law_of_thermodynamics#Specific_heat

and derive conclusions for the balance between the two subsystems. And we do this even without the precise analysis of each of the two subsystems.

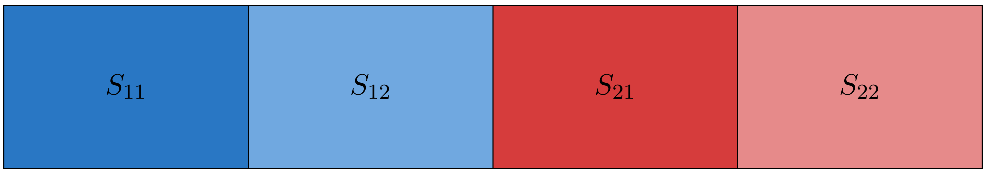
First of all, we are interested in the *second law*, and therefore we consider an isolated system which is decomposed into two parts with mean temperatures T_1 and T_2, deferring the question of the exact definition of what we mean by the mean value. And further we take – according to what has been said so far – that Eqn. 2.9 and 2.10 or also 2.11 express a first reasonable estimate for the entropy increase of the system. The following figure sketches the system with its two subsystems S_1 and S_2:



S_1	S_2

Table 2.2: Two-part system

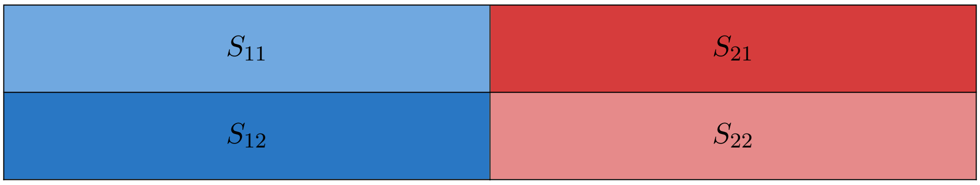
The assumption for the physical meaning of this decomposition is that the global heat exchange has also a directly global aspect[13] and points to the existence of global relations and is not alone the spatial expansion of local relations taking place in time. If we even assume that the decomposition of the system into the subsystems S_1 and S_2 is nothing else than the decomposition of all relations into those (global) across the two subsystems and all remaining (local) relations within each subsystem, then the decomposition implies at the same time that from the isolation of the whole system with respect to the determination of entropy also the isolation of each subsystem S_1 ,S_2 follows, insofar as its external relations are already completely captured by the entropy between both systems and the isolation of the system as a whole. The continuation of the decomposition of each subsystem in its turn into subsystems



S_{11}	S_{12}	S_{21}	S_{22}

Table 2.3: Example 1: Continued division of the two-part system

or e.g.



S_{11}	S_{21}
S_{12}	S_{22}

Table 2.4: Example 2: Continued division of the two-part system

therefore also implies the continuation of the entropy determination on the basis of the entropy autonomy of each subsystem in such a way that the entropy – besides an external measure – can be traced back to the measure of exclusively internal relations, which for each system is independent of its environment.

---

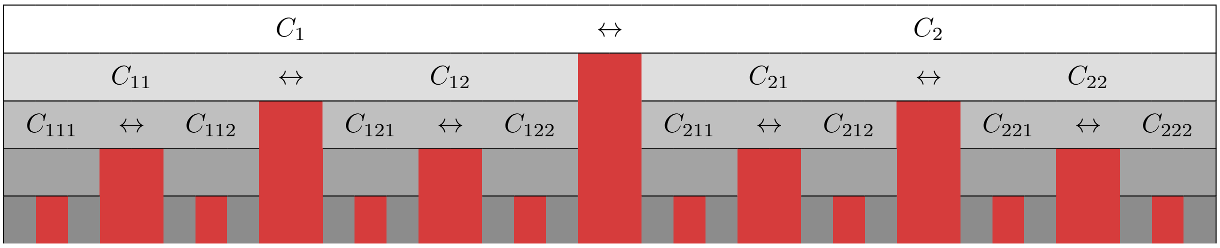
[13] Cf. section 5.2.2 on the contrast of relations and particles in three-dimensional space

Overall, this results in a hierarchical recursive relationship for the entropy determination, which differs from the linear-neighborhood relationship between the cells of the previous section. This can be illustrated, for example, as follows for eight cells C_1 to C_8:

C_1	$\leftrightarrow$	C_2	$\leftrightarrow$	C_3	$\leftrightarrow$	C_4	$\leftrightarrow$	C_5	$\leftrightarrow$	C_5	$\leftrightarrow$	C_7	$\leftrightarrow$	C_8

Table 2.5: Linear-neighborhood (extensive) relationship

Illustrating hierarchical recursive exchange, on the other hand, requires a representation taking into account different scales, which are shown line by line from top to bottom down the scale in the following figure:



			C_1				$\leftrightarrow$				C_2			
	C_{11}		$\leftrightarrow$		C_{12}		🟥		C_{21}		$\leftrightarrow$		C_{22}	
C_{111}	$\leftrightarrow$	C_{112}	🟥	C_{121}	$\leftrightarrow$	C_{122}	🟥	C_{211}	$\leftrightarrow$	C_{212}	🟥	C_{221}	$\leftrightarrow$	C_{222}

Table 2.6: Hierarchical (intensive) correlation

The red colored areas in the table represent neighborhood relationships between the two neighbors to the left and to the right, whose boundary heat flux is disregarded at the respective scale because this already occurs upscale.

First, it can be stated that the basic monotonicity $\Delta S \geq 0$ of Eqn. 2.19 and 2.26 of the linear relation follows directly in the case of the hierarchical relation, since each subsystem – as a two-part system – makes a non-negative contribution to the entropy according to Eq. 2.11, if one additionally assumes that the different contributions to the total entropy are merged completely or partially – in principle summarily, anyway. Each scale represents an estimate to be added to the total entropy.

However, the question arises about the relation of both correlations. For the linear relation, entropy is a sum of neighborly exchange entropies of parity subsystems – with the exception of those of the periphery. The hierarchical relation sums exchange entropies of neighboring subsystems as well, but because of the hierarchy, subsystems of all possible sizes are included in the sum. Therefore, it makes sense to anticipate section 3.3.3 about the complexity of partition hierarchies and locality and to demonstrate that e.g. the Shannon entropy is identical for both perspectives, which also puts the theses of locality and globality on an equal footing. For instance, let us start from the hierarchical relation according to table 2.6 and abstractly describe the relations between the subsystems C_1 and C_2 by the Shannon entropy

$$S(C_1, C_2) := -p_1 \log(p_1) - p_2 \log(p_2)$$

with the possibly unequal weights p_1 and p_2, for which $p_1 + p_2 = 1$ holds. In the same way we can describe the relations between the subsystems C_{11} and C_{12} and also those between C_{21} and C_{22}:

$$S(C_{11}, C_{12}) := p_1(-p_{11} \log(p_{11}) - p_{12} \log(p_{12}))$$
$$S(C_{21}, C_{22}) := p_2(-p_{21} \log(p_{21}) - p_{22} \log(p_{22}))$$

Then, because of $p_{11} + p_{12} = 1$, we get:

$$\begin{gathered}
-p_1 \log(p_1) + p_1(-p_{11}\log(p_{11}) - p_{12}\log(p_{12})) = \\
-p_1(p_{11}+p_{12})\log(p_1) + p_1(-p_{11}\log(p_{11}) - p_{12}\log(p_{12})) = \\
-p_1 p_{11}\log(p_1 p_{11}) - p_1 p_{12}\log(p_1 p_{12})
\end{gathered}$$

Since the analogous holds for the subsystem C_2 and its subordinate subsystems, it follows overall because $p_1 \cdot p_{11}, p_1 \cdot p_{12}, p_2 \cdot p_{21}, p_2 \cdot p_{21}$ is the probability distribution for the subsystems C_{11}, C_{12}, C_{21} and C_{22}:

$$S(C_1, C_2) + p_1 S(C_{11}, C_{12}) + p_2 S(C_{21}, C_{22}) = S(C_{11}, C_{12}, C_{21}, C_{22}) \tag{2.34}$$

The relationship of the hierarchically arranged systems, represented by the left side of the above equation, is thus equal to the relationship of the linearly arranged systems, represented by the right side of the equation (cf. also Eqn. 3.13 and 3.15 f.).

2.2.3 Physically distinguished points

A consequence of the linear-hierarchical relation is as follows: Let

$$f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

be a possibly discontinuous but integrable function in the second argument. The first argument may express holistic dependence on an environment, so that the function values change, especially when the environment is increased. Suppose, for example, that physical meaning in terms of measurability is only sums (integrals $\int f(x, x^{'})\, dx^{'}$ and averages over intervals with possibly very short interval lengths ≥ 0) of the function values of f and, in addition, that according to the last section also each scale contributes to the summation of f, but on each scale always only one half around a selected point, as shown by the white and gray highlighted cells of the following figure:

Table 2.7: Asymmetrisierung durch Auswahl

In a continuous scale space, the one-sided selection for summation according to table 2.7 corresponds to a double integral

$$F(x) = \pm 1 \int_0^x \left(\int_0^{x^{'}} f(x^{'}, x^{''})\, dx^{''} \right) dx^{'} \tag{2.35}$$

for which the integral upper bound x of the outer integral has the meaning of a continuous reciprocal scale, and the inner integral represents the actual summation of the function values of f. The function F defined by the integral is the physical quantity derived from f in relation to the reference point 0. By the method of double summation all points $x^{'}$ are distinguished in relation to other points $x > x^{'}$, in a special way – more than all others – the point 0.

The opposite of the above distinction is averaging

$$F(x) = \frac{1}{x}\int_0^x f(x, x^{'})\, dx^{'} \tag{2.36}$$

which resembles a smoothing and reduces the highlighting of distinguished points in the interval $[0, x]$. For those functions which represent distributions without any distinguished points, i.e. the constant functions or the uniform distributions, the averaging reproduces constant functions. Thus, if $f \equiv 1$, then:

$$F(x) = \frac{1}{x}\int_0^x 1\, dx^{'} = 1$$

Another special case is when the award according to Eqn. 2.35 leads to the same result as the smoothing according to Eqn. 2.36. Then, the inner integral in Eqn. 2.35 can be replaced by $x^{'}F(x^{'})$ according to Eqn. 2.36. We then receive

$$F(x) = \pm \int_0^x x^{'}F(x^{'})\, dx^{'}$$

and by derivation

$$F^{'}(x) = \pm xF(x)$$

which finally is solved by:

$$F(x) = \frac{1}{\sqrt{2\pi}} e^{\pm\frac{x^2}{2}} \tag{2.37}$$

The sign in the exponent of F can be positive or negative. In both cases F describes a physically distinguished point, in the first case with the maximum of the normal distribution and in the second case with a minimum.

Thus, the normal distribution is equally related to the asymmetry as to the symmetry of space. In its symmetric property it is the uniform distribution of the infinite extensive space.[14]

2.3 Expansion, contraction and time reversal

2.3.1 The heat equation

Connection with entropy increase

From the *continuity equation*[15], *Fourier's law*[16] and the definition of heat capacity (cf. in particular *heat flow in a uniform rod*[17]) follows the heat equation:

$$\frac{\partial T}{\partial t} = \alpha \Delta T \tag{2.38}$$

The definition of the heat capacity c, or its relationship to the density ρ and temporal changes in heat and temperature at a location, is:

---

[14] Cf. section 3.5.4

[15] https://en.wikipedia.org/wiki/Continuity_equation#Energy_and_heat

[16] https://en.wikipedia.org/wiki/Thermal_conduction#Fourier's_law

[17] https://en.wikipedia.org/wiki/Heat_equation#Heat_flow_in_a_uniform_rod

$$\frac{\delta Q}{dt} = c\rho \frac{\partial T}{\partial t} \tag{2.39}$$

Thus, the heat equation can also be derived from Eqn. 2.24, in which the continuity equation and Fourier's law have already been processed: First, it follows from Eqn. 2.24:

$$\frac{\delta Q(\mathbf{r})}{dt} = k \cdot \Delta T$$

Further, because δQ in Eqn. 2.39 is equal to the net heat flux per unit volume, the left-hand side of this equation is identical to the left-hand side $\delta Q/dt$ of the Equation above, so the two right-hand sides also agree:

$$c\rho \frac{\partial T}{\partial t} = k\Delta T$$

With $\alpha := k/c\rho$, the heat equation 2.38 follows.

In this derivation, the peripheral boundary condition 2.17, which implies the conservation of energy by isolation according to Eqn. 2.20, is not assumed, rather only the local relation $\delta Q_i = \delta q_i - \delta q_{i-1}$ of the *continuity equation* in the form of Eqn. 2.15.

In summary: The heat equation 2.38 summarizes three of the four conditions for the entropy increase according to Eq. 2.26. It lacks the condition of isolation or global conservation of energy.

Interpretation of the heat equation

In the interpretation of the heat equation given *here*[18], it describes, from the perspective of a point $\mathbf{r}$ in space, the infinitesimal process of temperature change at $\mathbf{r}$ over time. The process is decisively determined by the difference

$$-[T(\mathbf{r}) - T(U(\mathbf{r}))]$$

between the temperature $T(\mathbf{r})$ at this point and the mean temperature $T(U(\mathbf{r}))$ of its environment $U(\mathbf{r})$, which does not contain $\mathbf{r}$ itself. This follows, although very simplified, but quickly also directly from *Fourier's law* and Eqn. 2.23. First according to Fourier's law and Eqn. 2.21, for a cube ijk with $\Delta x = \Delta y = \Delta z = \Delta l$ the following equations

$$\begin{aligned}
\delta q_{ijk,x} - \delta q_{(i-1)jk,x} &= k(T_{(i+1)jk} - T_{ijk})\Delta l \Delta t - k(T_{ijk} - T_{(i-1)jk})\Delta l \Delta t \\
\delta q_{ijk,y} - \delta q_{i(j-1)k,y} &= k(T_{i(j+1)k} - T_{ijk})\Delta l \Delta t - k(T_{ijk} - T_{i(j-1)k})\Delta l \Delta t \\
\delta q_{ijk,z} - \delta q_{ij(k-1,z)} &= k(T_{ij(k+1)} - T_{ijk})\Delta l \Delta t - k(T_{ijk} - T_{ij(k-1)})\Delta l \Delta t
\end{aligned}$$

hold and thus also according to eqn. 2.23:

$$\delta Q_{ijk} = -6k\Delta l \Delta t \left(T_{ijk} - \frac{T_{(i+1)jk} + T_{(i-1)jk} + T_{i(j+1)k} + T_{i(j-1)k} + T_{ij(k+1)} + T_{ij(k-1)}}{6} \right) \tag{2.40}$$

---

[18] https://en.wikipedia.org/wiki/Heat_equation#Interpretation

In this equation all temperatures are mean values. The fraction is therefore an average of the temperature averages of the 6 cubes surrounding the cube ijk, and it is thus the temperature average of the environment U_{ijk} of the cube ijk formed by the 6 cubes, which itself is not contained in U_{ijk}. To make it clearer how the result is related to the spatial order and the temperature curvature ΔT, we again consider only a one-dimensional space with three cells, an inner cell with temperature T and its neighboring left and right cells with temperatures T_l and T_r:

T_l	T	T_r

Table 2.8: System of three cells

The mean temperature of the environment is $1/2(T_l + T_r)$, and for its difference from the internal temperature T

$$\frac{T_l + T_r}{2} - T = \frac{1}{2}[(T_r - T) - (T - T_l)]$$

applies. Finally, the difference in the bracket on the right side of this equation corresponds in the limit of small cells with the curvature $\partial^2 T/\partial x^2$, as it is shown in more detail by the calculation of the net heat flux in 2.2.1.

If the result is additionally considered against the background of energy conservation, e.g. in the simplified form according to Eqn. 2.32, i.e. under the condition of a constant global temperature mean, then the two theorems establish the rules of a zero-sum game applicable to spatial neighbors, while the heat equation describes the game evolution whose outstanding feature is that each point bows to the majority decision of its surrounding collective in such a way that local extrema of the temperature distribution are degraded and the points of space distinguished by them increasingly fade away. Altogether, the heat equation with its interpretation of infinitesimal averaging is the differential interpretation of a global smoothing and propagation process, as shown by the fundamental solution of the equation, which for the one-dimensional case is

$$T_F(x, t - t_0) \sim \frac{1}{\sqrt{4\pi\alpha(t - t_0)}} e^{-\dfrac{x^2}{4\alpha(t - t_0)}} \tag{2.41}$$

which is a normal distribution with monotonically increasing variance $\sigma^2 = 2\alpha(t - t_0)$ as a function of time.

Time reversal and time symmetry

The fundamental solution shows that it is not a solution of the equation for negative co-efficients α. Now this change of sign is equivalent to the time reversal (see below). The heat equation thus becomes a *backward parabolic differential equation*[19]. As the just referenced article in Wikipedia shows, the equation is even considered as an improperly posed problem in that case:

> *An initial/boundary-value problem for a backward parabolic PDE is usually not well-posed (solutions often grow unbounded in finite time, or even fail to exist).*[7]

---

[19] https://en.wikipedia.org/wiki/Parabolic_partial_differential_equation#Backward_parabolic_equation

In doing so, the article refers to the definition of the correctly posed problem given by the French mathematician *Jacques Hadamard*, as the following quote shows:

> *The mathematical term well-posed problem stems from a definition given by 20th-century French mathematician Jacques Hadamard. He believed that mathematical models of physical phenomena should have the properties that: 1. a solution exists, 2. the solution is unique, 3. the solution's behaviour changes continuously with the initial conditions.*[9]

This article gives as an example of an incorrectly posed problem:

> *Problems that are not well-posed in the sense of Hadamard are termed ill-posed. Inverse problems are often ill-posed. For example, the inverse heat equation, deducing a previous distribution of temperature from final data, is not well-posed in that the solution is highly sensitive to changes in the final data.*[9]

In any case, this reveals with clarity that the equation with negative coefficients describes completely different realities. In fact, the change turns all properties of the equation with positive coefficient upside down. First of all, this includes the reversal of the smoothing into an unrestrained spreading of local temperature differences. Thus, the equation has basically the tendency to increasingly mark out singular spatial points, those with local temperature maximum or minimum, in front of the points of their respective surroundings. At the same time, however, neighboring and non-neighboring extrema compete with each other in a complex way, so that solutions of the equation will generally be unstable depending on their initial values. But also here the equation describes the course of a zero-sum game, but this time with actors exclusively competing for the globally constant resource energy.

Thus, while positive coefficients level differences by smoothing, avoid or reduce competition, and imply stable solutions, negative coefficients favor the opposite, the growth of existing differences, (apparent) spontaneity, the emergence of competition, and the state of permanent instability associated with the extremely sensitive dependence on initial values.

The change of sign of the coefficient α in the heat equation 2.38 is equivalent to the reversal of time $dt \to -dt$, because only the temporal change of the temperature depends on the direction of time:

$$\frac{\partial T(\mathbf{r}, t-t_0)}{-\partial t} = \alpha \Delta T(\mathbf{r}, t-t_0) \Leftrightarrow \frac{\partial T(\mathbf{r}, t-t_0)}{\partial t} = -\alpha \Delta T(\mathbf{r}, t-t_0) \tag{2.42}$$

$$\frac{\partial T(\mathbf{r}, t-t_0)}{-\partial t} = -\alpha \Delta T(\mathbf{r}, t-t_0) \Leftrightarrow \frac{\partial T(\mathbf{r}, t-t_0)}{\partial t} = \alpha \Delta T(\mathbf{r}, t-t_0) \tag{2.43}$$

According to the above, the heat equation thus describes – with fixed sign of α – a completely different behavior for the evolution into the future as for that into the past and in this sense breaks the symmetry of time. While, for $\alpha \geq 0$, the temperature distribution locally as well as globally always approaches the unambiguous target of the uniform distribution or the linear distribution, the development in backward time direction is neither determined nor stable insofar as its development depends on smallest fluctuations in the initial values.

Time reversal and reversibility

However, the instability and sensitive dependence of the backward state development on all details of the state is not equivalent to an irreversibility in the sense of a fundamental impossibility to reconstruct past states unambiguously on the basis of the current state. The irreversibility is seemingly obvious because of the smoothing, for the completed smoothing in the form of the uniform distribution or the linear distribution is undoubtedly not reversible indicating that already the beginning of smoothing might exclude the possibility of a definite reconstruction. On the other hand, the fundamental solution 2.41 gives an example of the fact that the uniform distribution is not reached in finite time, and in its case the distributions of the past for the period (t_0, t) are unambiguously derived from the distribution at time t. Also, in this interval it is actually a solution of the heat equation with negative coefficient α and backward running time, because for the fundamental solution

$$T_F(\mathbf{r}, t - t_0, \alpha) = T_F(\mathbf{r}, t_0 - t, -\alpha)$$

holds. Reversing the time direction $dt \to -dt$, because T_F satisfies the heat equation in positive time direction with positive coefficient, results in:

$$\frac{\partial T_F(\mathbf{r}, t_0 - t, -\alpha)}{-\partial t} = \frac{\partial T_F(\mathbf{r}, t - t_0, \alpha)}{-\partial t} = -\alpha \Delta T_F(\mathbf{r}, t - t_0, \alpha) = -\alpha \Delta T_F(\mathbf{r}, t_0 - t, -\alpha) \quad (2.44)$$

Reversibility means that the state function does not suffer any loss of information by the update in time, i.e. structures are smoothed but do not disappear – in finite time. The fundamental solution is a perfect example of a smoothing without structural change, since the update in time is nothing else than a reversible and norm-preserving transformation[20] in space, which assigns the distribution $T_{\sigma(t)}$ to the initial temperature distribution T_0 for each time $t > t_0$. For with $f(x) := e^{(-1/2)x^2}$ Eqn. 2.41 becomes

$$T(x, t - t_0) \sim \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

where in the dependence $\sim (1/\sigma) f(x/\sigma)$ the norm conservation is shown, with which the abscissa is stretched by the increasing factor σ and the ordinate values are compressed in return by the same factor. Incidentally, as a consequence of this, the integral over the temperature and thus also the temperature mean is time-independent, which follows more precisely from *Integration by substitution* as follows:

$$\int_{-\infty}^{+\infty} T(x)\, dx \sim \frac{1}{\sigma} \int_{-\infty}^{+\infty} f\left(\frac{x}{\sigma}\right) dx = \int_{-\infty}^{+\infty} f(x)\, dx \quad (2.45)$$

At this point, it may also be noted that the fundamental solution T_F has no finite entropy because T_F is a solution of the heat equation 2.38, for which the coefficient α and hence the density ρ are assumed to be constant in space (and time). Based on the heat equation, we also derived Eq. 2.31, from which, for the entropy of the fundamental solution follows:

$$S = \rho c \int \log\left(\frac{T}{T_R}\right) dx \sim \int_{-\infty}^{+\infty} x^2\, dx$$

Let us return to the reversibility of the fundamental solution, which is obviously related to the fact that the scale transformation on the abscissa is unrestricted by the possibility of unlimited expansion in infinite extensive space, which does not apply to spatially finite systems by definition. Here, let us say, squeezing is unavoidable, i.e. the expansion does not take place in extensive

---

[20] Cf. section 3.5.5

space but in intensive space. To examine this more closely, we consider the temperature change dT after the expiration of the time dt, as it results according to the heat equation:

$$dT_+(\mathbf{r}, t) = \alpha \Delta T(\mathbf{r}, t) dt$$

Now we let time run backwards, i.e. we evolve the changed temperatures backwards according to the heat equation with the negative coefficient $-\alpha$. We obtain:

$$dT_-(\mathbf{r}, t + dt) = -\alpha \Delta T(\mathbf{r}, t + dt) dt$$

If the backward heat equation allows to revise the infinitesimal changes of the forward equation, then the error

$$err(\mathbf{r}, t) := dT_+(t) + dT_-(t + dt)$$

must be sufficiently small. The following results from the two equations

$$err(\mathbf{r}, t) = -\alpha(\Delta T(\mathbf{r}, t + dt) - \Delta T(\mathbf{r}, t)) dt$$

and with it:

$$err(\mathbf{r}, t) = -\alpha \frac{\partial \Delta T}{\partial t} \cdot dt^2 = -\alpha \frac{\partial^2 T}{\partial t^2} \cdot dt^2 \tag{2.46}$$

If the time derivative $\partial \Delta T / \partial t$ or equivalently the second derivative $\partial^2 T / \partial t^2$ exists and is bounded, then the error is a second order differential and therefore very small relative to the temperature change itself, so that reconstruction is possible even over finite time intervals.

However, also the conditions of reversibility become clear, which consist in the fact that the spatial as well as the temporal temperature changes must be sufficiently smooth also on arbitrarily small scales. If this is the case, then the local process of the heat equation possibly corresponds to a local scale transformation which is reversible with sufficient accuracy. However, if temperature fluctuations do not break down even in the limit of small space or time scales, then the above equation indicates the magnitude of the scale-dependent error. Smooth, differentiable temperatures are then only temperature averages on finite scales. The mean value interpretation of the heat equation presented in section 2.3.1 shows that the process of heat conduction based on averaging in an infinitesimal environment of a point $\mathbf{r}$ in space can be well defined even for non-differentiable temperature distributions, as described in that section.

Self-reference and the singularity of the fundamental solution

The fact that the reconstruction of the fundamental solution at time t_0 ends in a singularity has a deep and in any case no accidental meaning, which can be seen better by looking at *diffusion*[21] of particles, because this, unlike the heat diffusion, can also be visible in everyday life in the true sense of the word. The *diffusion equation*[22] is also identical in its simple form with the heat equation, so that diffusion can actually be used to illustrate heat diffusion as well. In particular, the phenomenon of diffusion in the coffee cup, which we all know, can be used here, which we observe when coffee and poured milk or cream mix with each other. After some time, we perceive a homogeneous, i.e. evenly distributed, mixing state. If we would film the diffusion process with a camera and let the film run backwards, out of – the assumed – uniform distribution unpredictable colorations would arise, reflecting concentration clusters of the coffee or cream particles, of which some – which, again, is practically not determined – increase in intensity. If one assumes that the real process completely follows the diffusion equation, then

---

[21] https://en.wikipedia.org/wiki/Diffusion

[22] https://en.wikipedia.org/wiki/Diffusion_equation

the said indeterminacies of the backward evolution are, as mentioned above, however, only a consequence of incomplete knowledge in detail. The time reversal singularity at $t = t_0$ shows the limits of the model describing diffusion as a self-referential process, self-referential because the time evolution of the state function, in the case of heat diffusion the temperature, depends exclusively on the state function itself, namely its curvature ΔT. In other words, the system determines the evolution of its state itself, in positive as well as negative time direction. This self-reference of the system is now interrupted at time $t = t_0$, the time of the unification with another system , so that the diffusion equation at t_0 can also have no solution, since it only describes the self-referential development of the system, and this therefore – also in backward time – would only remain with itself, but actually interacts with the environment. In this respect, the singularity of the fundamental solution is an expression of the fact that the self-reference of the system comes to a conclusion, i.e. the system can no longer describe itself.

The continuity equation

At the beginning of section 2.3.1, we pointed out that the heat equation follows from the continuity equation for heat energy, Fourier's law and the connection of the local temperature change on one side and the heat capacity and density on the other side. This connection allows to replace the heat energy in the continuity equation – for heat energy – by the temperature, so that one can expect the heat equation to be a continuity equation – for temperature – as well. To see this clearly, let us consider the *general continuity equation*[23] for a quantity ρ that is in differential form and with the definition of the flux density

$$\mathbf{j} = \rho \mathbf{u} \tag{2.47}$$

as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \cdot \mathbf{u}) = \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = \sigma \tag{2.48}$$

In it u is a velocity field, according to which the intensive quantity ρ, which in our imagination is often a density, moves dependent on location and time. If ρ is locally conserved, then $\sigma \equiv 0$. According to the *Gaussian divergence theorem*[24] the integral form is also valid. Specifically, for a locally conserved quantity ρ holds:

$$\frac{d}{dt} \int \rho \, dV + \oint \mathbf{j} \cdot d\mathbf{S} = 0 \tag{2.49}$$

In this equation, the first integral is a volume integral over the conserved quantity in a region G and the second is a surface integral over the scalar products $\mathbf{j} \cdot d\mathbf{S}$ on the surface of the region. The surface integral calculates the total current flowing out of area G through the surface minus the total current flowing in.

We can now put the heat equation 2.38 into the following form

$$\frac{\partial T}{\partial t} + \nabla \cdot \left(T \cdot \left(-\alpha \frac{\nabla T}{T} \right) \right) = 0 \tag{2.50}$$

from which, by comparison with the general form of the continuity equation 2.48, it follows immediately that the heat equation is also a continuity equation with density $\rho = T$ and velocity

---

[23] https://en.wikipedia.org/wiki/Continuity_equation#General_equation

[24] https://en.wikipedia.org/wiki/Divergence_theorem

field:

$$\mathbf{u} = -\alpha \frac{\nabla T}{T}$$

Eqn. 2.49 makes it possible to identify a simple and sufficient criterion for the locally conserved quantity to be also a globally conserved quantity. Namely, this is certainly the case if $\mathbf{j} \cdot d\mathbf{S} = 0$ holds on the whole surface of the domain G, i.e. if the current everywhere along the surface disappears or at least runs parallel to it, and the system thus has no exchange with the environment of G. For a thermodynamic system, we would say in that case that the system is isolated, which corresponds to the definition of the open rod in section 2.2.1. The integral form of the heat equation 2.50 is then

$$\frac{d}{dt} \int T \, dV = 0$$

from which, as in Eqn. 2.32, the time constancy of the temperature mean value $\frac{1}{V} \int T \, dV$ in the defined area G results.

Another consequence of this section is that not only the heat equation and the diffusion equation, but also the continuity equation is related to breaking the symmetry of time.

Single vs. double stream

As the previous sections show, time symmetry is broken by the heat equation and the diffusion equation. In contrast, space symmetry remains unaffected because the equations are invariant to the reversal of the direction in space as a result of the second derivative $\partial^2/\partial x^2$. The consequences of time reversal are reflected by the equivalences 2.42 and 2.43, which go back to the fact that the time derivative is of first order only.

The continuity equation 2.15 represents both relations, with respect to time and space, at the same time. First, reversing directions in space leaves the continuity equation

$$\delta Q_i = (-\delta q_{i-1}) - (-\delta q_i) = \delta q_i - \delta q_{i-1} \tag{2.51}$$

and thus δQ_i unchanged. The second order derivative mentioned above is expressed in Eqn. 2.51 through the sign reversal of the boundary heat fluxes $\delta q_i \to -\delta q_i$ in conjunction with the cell order reversal represented by the index i. The fundamental possibility for this reflects the symmetry of space, which does not favor either direction and leaves room for directional definition. Once one direction is chosen as preferred, the sign of the boundary heat fluxes δq_i is uniquely determined by the second law $\mathrm{sgn}(\delta q_i) = \mathrm{sgn}(T_{i+1} - T_i)$ (Cf. Eqn. 2.18). On the other hand, as already mentioned, the sign of the *net* heat flux δQ_i and thus that of the coefficient α in the heat equation is independent of the said choice and therefore exclusively determined by the second law, so that overall the broken symmetry of time is solely a consequence of the second law and its universal validity.

Obviously, the time symmetry would be fulfilled, if there would be not only one, but beside this a second stream, so that besides the universality of the second law at the same time its inversion could manifest itself. This is obvious in so far as – also in the classical particle model – the heat stream and in our conception above all the particle stream along a space axis actually seems to consist of two streams, since particles and impulses move in the one as well as in the other direction. However, the prerequisite for the derivation of the heat equation and the diffusion equation, but also for the definition of the more general continuity equation, is always a singular current, which combines the two parts of the – according to our conception existing – double stream. But the fact that it is indeed a single stream only follows from the indistinguishability of the particles. Because the distinction of particles with opposite velocities presupposes also their distinguishability.

A double stream would first of all have the potential to preserve the symmetry of time and therefore also the potential to allow expansion and contraction at the same time in only one direction of time.

2.3.2 The Schrödinger equation

Time symmetry of the double stream

In a summary of the previous sections we can say the following: The fundamental solution 2.41 and more generally the mean value interpretation of the heat equation in section 2.3.1 shows that the equation describes an expansion process. Thus, the heat equation and with it the diffusion equation – in positive time direction and with positive coefficient α – cannot represent the reverse process of contraction in principle. In other words, in one time direction it can describe either only the expansive or only the contractive process.

Since now both expansion and contraction are obviously part of reality in positive time direction, the question remains for an equation which is able to represent both aspects of reality at the same time. Such an equation can possibly be symmetrical also in time, because with time reversal it then maps expansion and contraction likewise.

It turns out that it is the *Schrödinger equation*[25] which reflects the dualism of both realities of expansion and contraction – in one time direction – and at the same time elevates the symmetry of time even to a principle. The Schrödinger equation of a particle with the potential $\phi(r,t)$ is in its general complex representation

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi(\mathbf{r},t) + \phi(\mathbf{r},t)\psi(\mathbf{r},t) \tag{2.52}$$

which is equivalent to $\frac{\partial\psi}{\partial t} = i\frac{\hbar}{2m}\Delta\psi - i\frac{\phi}{\hbar}\psi$, from which, if we convert ψ into its real ψ_a and imaginary part ψ_b so that $\psi = \psi_a + i\cdot\psi_b$ holds, it follows:

$$\frac{\partial\psi_a}{\partial t} + i\frac{\partial\psi_b}{\partial t} = -\frac{\hbar}{2m}\Delta\psi_b + i\frac{\hbar}{2m}\Delta\psi_a - i\frac{\phi}{\hbar} + \frac{\phi}{\hbar}\psi_b$$

This equation now is equivalent to the two coupled differential equations

$$\frac{\partial\psi_a}{\partial t} = \left(-\frac{\hbar}{2m}\Delta + \frac{\phi}{\hbar}\right)\psi_b \tag{2.53}$$

$$\frac{\partial\psi_b}{\partial t} = \left(+\frac{\hbar}{2m}\Delta - \frac{\phi}{\hbar}\right)\psi_a \tag{2.54}$$

for the two (twin) functions ψ_a and ψ_b. The comparison of both equations for vanishing potential $\phi \equiv 0$ with the heat equation 2.38 shows that, in contrast to the latter, the Schrödinger equation of the free particle is dualistic and presumably unites both sides of reality – the expansion and the contraction, the collectivistic subordination and the individualistic distinction as well as the accompanying stability and instability with the coefficients $\alpha = +\hbar/2m$ and $\alpha = -\hbar/2m$.

---

[25] https://en.wikipedia.org/wiki/Schrödinger_equation

As in the case of the heat equation[26], the time reversal $dt \to -dt$ changes the sign in both equations only on the left side and thus fundamentally changes the equation system. But this change is equivalent to the exchange of the two – in the system of equations – anti-symmetric functions ψ_a and ψ_b, so that the time reversal leaves the system structurally unchanged and both time directions are therefore of the same kind and in principle have the same development potential.

In particular, the two functions – unlike the heat conduction equation, the diffusion equation, and the continuity equation, each with only one (density) function – represent the double stream mentioned in the previous section for maintaining time symmetry. Each of the two streams by itself, but not their unity, favors one direction of time. Analogous to heat conduction[27] represents

$$-\Delta\psi_b = \frac{\dfrac{\partial\psi_b(x)}{\partial x} - \dfrac{\partial\psi_b(x+dx)}{\partial x}}{\partial x}$$

the relative difference of two boundary streams (momentum) at the boundaries of a cell C_n, as shown in Table 2.9 below. With respect to this cell, $\partial\psi_b(x)/\partial x$ is the value of an inflowing stream and $\partial\psi_b(x+dx)/\partial x$ is the value of an outflowing stream, so the difference calculates a net influx into cell C_n, which – like the boundary streams of heat conduction – can be positive or negative and can be registered in the cell as a change $\partial\psi_a/\partial t$ of ψ_a, if the potential satisfies $\phi \equiv 0$, i.e. if it is a free particle.

C_{n-1}	C_n	C_{n+1}
$\frac{\partial\psi_b(x)}{\partial x} \to$	$\frac{\partial\psi_b(x+dx)}{\partial x} \to$	$\dots$
	$\frac{\partial\psi_a}{\partial t}$	

Table 2.9: Boundary flux $\partial\psi_b/\partial x$ and net influx $\partial\psi_a/\partial t$ of a free particle

The second part 2.54 of the Schrödinger equation represents the same relationship but with respect to the reversed direction of space with an inflow $\partial\psi_a(x+dx)/\partial x$ into and an outflow $\partial\psi_a(x)/\partial x$ from cell C_n:

C_{n-1}	C_n	C_{n+1}
$\frac{\partial\psi_a(x)}{\partial x} \leftarrow$	$\frac{\partial\psi_a(x+dx)}{\partial x} \leftarrow$	$\dots$
	$\frac{\partial\psi_b}{\partial t}$	

Table 2.10: Boundary flux $\partial\psi_a/\partial x$ and net influx $\partial\psi_b/\partial t$ of a free particle

---

[26] Cf. the equivalences 2.42 and 2.43

[27] Cf. the magnitude of the net heat flow in section 2.2.1

Double stream vs. single probability stream

Multiplying Eqn. 2.53 by the function ψ_a and Eqn. 2.54 by ψ_b gives:

$$\frac{1}{2}\frac{\partial}{\partial t}{\psi_a}^2 = \psi_a\frac{\partial\psi_a}{\partial t} = -\frac{\hbar}{2m}\psi_a\Delta\psi_b + \frac{\phi}{\hbar}\psi_a\psi_b \tag{2.55}$$

$$\frac{1}{2}\frac{\partial}{\partial t}{\psi_b}^2 = \psi_b\frac{\partial\psi_b}{\partial t} = +\frac{\hbar}{2m}\psi_b\Delta\psi_a - \frac{\phi}{\hbar}\psi_a\psi_b \tag{2.56}$$

Adding both equations and multiplying by 2 results in

$$\frac{\partial}{\partial t}\left(\psi_a^2 + \psi_b^2\right) = \frac{\hbar}{m}(\psi_b\Delta\psi_a - \psi_a\Delta\psi_b) \tag{2.57}$$

which is consistent with the *quantum mechanical continuity equation*[28]. Since the magnitude square $\psi_a^2+\psi_b^2$ of the wave function in quantum mechanics is considered as the probability density function of finding the particle in space, the left side of the above equation is the temporal change of this probability and the right side is therefore the associated probability stream, which can be interpreted as such in particular because it does not explicitly depend on the potential ϕ, and as a consequence the density change on the left side is exclusively caused by the stream.

In the equation, the relation between time reversal and the anti-symmetry of the functions ψ_a and ψ_b holds unchanged, i.e., the time reversal is equivalent to the interchange of the two functions, so that also the probability stream can evolve in both directions of time in the same way and has the potential for contractive as well as for expansive state evolutions in both directions, as also the discussion about the question *Is it possible for Δx of any free particle wave packet to be decreasing at any time?*[29] on *StackExchange* shows. Namely, it is shown there that while the momentum variance remains unchanged, the variance Δx^2 of a Gaussian wave packet can fall or grow in time t according to:

$$\Delta x^2 = \frac{\tau^2 - t^2}{\tau^2}$$

So also the free particle has the potential for contraction and equally for expansion, so that in particular the contraction is not necessarily the consequence of a continuing potential, which is analogous to a collective of classical free particles, which in principle can be in the state of contraction as well as in the state of expansion.

---

[28] https://en.wikipedia.org/wiki/Continuity_equation#Quantum_mechanics

[29] https://physics.stackexchange.com/questions/54534/is-it-possible-for-delta-x-sigma-x-of-any-free-particle-wave-packet-to-b?noredirect=1&lq=1

Chapter 3

Fundamentals of entropy

3.1 Complexity, diversity and real size

The interpretation of the *second law* and of entropy in general was early connected with the idea of molecular chaos[1], in which the particles move with great variety relative to each other or relative to a selected velocity. The increase of entropy is in this conception correspondingly the increase of this variety of motion. However, one must separate two things, i.e.

1. The global chaos of (relative) velocities
2. The distribution of such chaos in the position space

The global chaos is closely connected with the law of conservation of energy, which sets a cumulative limit to the range of relative velocities. If one neglects, as in an ideal gas, the potential energy of the particles, then the energy is even a direct measure for the global basically possible variety of velocities. But this also means that the kinetic energy itself does not force the variety, but only determines the size of the space for it by its relationship with the velocity variance. Possible and compatible with any energy are also statistically correlated states of motion with only a few relative velocities. The increase of the global velocity variety from a narrow to a broad spectrum is therefore a balancing, spreading development in the velocity or momentum space, just as the chaos distribution expresses the balancing and the spreading of this variety of motion in real space.

It is intuitively obvious to associate such diversity with complexity. More precisely, complexity is characterized by two features. First, its system size, i.e. the number of its parts, which conversely means that a small system with only a few parts cannot be complex. Second, its symmetry or asymmetry, i.e. the number of its broken symmetries, which means that a system is not necessarily complex by size alone, but only if the size is confirmed by the variety of broken symmetries. So there may be many parts, for example, but because of their sameness or similarity they increase the complexity only slightly or not at all. Conversely, it follows from this that complexity is the real size of the system, so that the complexity finally depends on only one parameter, this real size, which explains the simplicity of Boltzmann's formula 1.2:

$$S = k_B \log(W)$$

The fact that this complexity is not the number W itself, but its logarithm, does not contradict this, since after what has been said complexity only needs to grow monotonically with W. There are two possible interpretations for the formula:

---

[1] Cf. section 5.1.1 about the Loschmidt paradox

Counting principle: In a special interpretation for dynamical systems with a state space of possible states and e.g. two subsystems with W_1 and W_2 states, respectively, $W = W_1 \cdot W_2$ is the number of possible state combinations of the composite system and thus the number of all states of this system when the states of the two subsystems are independent. According to this counting principle already presented in the introduction, the complexity is due to

$$\log(W_1 \cdot W_2) = \log(W_1) + \log(W_2)$$

additive. Nevertheless, it is not physically extensive, as will be discussed in more detail in section 3.5.2. Also, the problem of this interpretation is already addressed in the introduction.

Description complexity: In the interpretation used in this work $\log(W)$ is the shortest description length of a system with W parts, as explained in section 3.3 dealing with the complexity of namespaces, which have no other purpose than to allow the abstract distinction of the W parts of a whole with the least possible effort.

The simplest example for a system, which has even unlimited size and nevertheless seems simple to us, is the real space. The reason for this simplicity is its *geometric symmetry*[2], especially the translation symmetry, but also rotational symmetry. Symmetry is a form of repetition, which – for description – allows the summary of parts or, in other words, the abbreviated description of parts with reference to a representing part, as for example in the natural numbers the successor relation $n \to n+1$ is an ever-repeating relation. However, despite this simplicity and symmetry, there is inherent asymmetry and concomitant complexity in extensive space and equally in intensive space, which is one of the key messages of the present work.

This fundamental nature of complexity as a measure of true size is well known from reality. Serial production, for example, is inexpensive because its manufacturing processes are always repeated in the same way and, conversely, individual production is cost-intensive because each part to be manufactured requires different manufacturing processes. Comparable to this is the routine, which is simple because of its repetition, while deviation from the routine or, for example, from a familiar organizational process is associated with complexity and with corresponding costs and not infrequently also with rejection and errors. As another example, sorting into collectives of parts each of the same kind has the purpose of reducing complexity, in that for each collective the description of a whole of parts each of the same kind is short and therefore simple. The description is short, because not every part has to be connected with a precise location, but a common, i.e. summarizing and fuzzy and therefore short specification of the location is sufficient for all parts. The fact that for the sorting (entropy reduction) costs have arisen before is another matter.

The last example of sorting, however, reveals another general property of complexity or its opposite, simplicity. For the sorting of representational things in physical space is the bringing together of at least two categories of order, of which space itself is one. Each category, if it is finite for simplicity, has a number of different characteristics. In computer science, the category would be denoted by the key and the features by the values for each category. So there is the category *space* for example with the characteristics *left* and *right* and for example the category *color* with the characteristics *white* and *black*. The sorting consists of merging e.g. the features *left* and *white* or *right* and *black*. The consequence of the merging is that the feature, e.g. *white*, of a thing of one category can be read from the feature *left* of the other category, so that the distinction of black and white can just as well be made by the distinction between left and right, but also – and this is essential for entropy – conversely the spatial distinctness becomes possible by that of the other category. Thus, for example, we make location indications even always with reference to objects in the real space like e.g. *beside or behind the tree* or *between the two clouds*,

---

[2] https://en.wikipedia.org/wiki/Symmetry_(geometry)

so that the matter and its distribution in space defines the reference system which only makes location indications possible.

A similar process takes place in a process of cognition, which also brings together the features of different concepts of order with the consequence that a complex inner state gives way to a simpler one.

In physics, the position space and the velocity or momentum space are the central ordering categories involved in all processes. If the particles of a physical system in the extreme case all have the same velocity and thus occupy only a small part of the velocity space then the description of the states of motion of all particles is possible in a simple and short way by specifying this one velocity and combining the particles to a single collective. If, on the other hand, the velocities differ from each other, then on the one hand the specification of only one velocity is inaccurate and on the other hand the specification of many velocities is complex and connected with corresponding expenditure.

In a more detailed consideration in the sense of the above mentioned combination of features of different order categories for the reduction of complexity, we first find out that in physics we have parts which we call particles, whose features are their positions and velocities or momenta, if we again assume otherwise indistinguishable particles for the sake of simplicity. The sorting and simplification mentioned above consists in assigning particles with the same velocity vector to a cell in position space. Thus the particles are sorted according to their states of motion, i.e. ordered, and at the same time the cells in space become distinguishable by the velocity of each of their particles. Finally, the cells in space are distinguishable even if the particles in a cell do not all have the same state of motion, but the state of motion averaged for each cell is its own. In this case the chaos in a cell is already large, since we assume different velocities of their respective particles, and it increases further with the velocity variance, which we finally connect in the local temperature equilibrium with the cell temperature. As long as the local temperatures (variances) are different, however, the particles are globally more or less correlated and spatially sorted according to their state of motion, so that the complexity of the total system has not yet reached its maximum.

3.2 Density entropy and motion entropy

The order categories merged in the last section are the position space on the one hand and the velocity or momentum space on the other hand. Now we notice that beside this merging there is another, but inauthentic merging. To see this, we deliberately assume indistinguishable particles and ignore their respective velocity characteristics. The distribution of particles in space basically corresponds to a sorting. On the other hand, there is nothing to sort because of their indistinguishability. Thus, two categories of order are not merged. Rather there is only one, space. The second aspect of the merging, which was emphasized as essential in the last section, nevertheless applies, namely the distinguishability of the spatial features, i.e. of cells of the space. There are cells with a particle and others without particles and, more generally, cells in space are distinguished by the number or density of their particles. Cells with high density are distinguished – in relation to their environment – before their neighboring cells.

Therefore, in order to interpret entropy, we have to distinguish between the two basic forms of entropy. The second law is usually associated only with the heat energy or the spatial variations of the temperature, as it is also done in chapter 2.2 about the second law. In this form, the law simply states that the local intensities of motion in an isolated system are continuously balanced with the progress of time on scales which are not too small. Together with the conservation of energy, this means that the local intensity of motion fluctuates spatially around a globally constant mean value and finally approaches this mean value.

As said, there are now beside the spatial fluctuations of the particle energy also those of the particle density (or the particle volume), which are connected likewise with a form of the entropy, which I would like to call density entropy in the following in contrast to the motion entropy. Intuitively, it can be assumed that the temporal monotonicity applies equally to the density entropy, if not, then to the sum of both entropies. Analogous to the entropy of motion, the density entropy is a measure for the averaged agreement of the local densities with the global density mean value, which reaches its maximum exactly in the case of the global equality of all local densities.

On the other hand, the primacy of the entropy of motion with respect to monotonicity is not accidental. For if, as said, it is to be assumed that also the density entropy does not decrease in an isolated system, then the balancing process of the motion entropy, which we call heat diffusion, is generally present and clearly perceptible, while the balancing process of the density entropy, which we call particle diffusion or simply diffusion, is clearly restricted by the existence especially of solids. In other words, particle diffusion is restricted and the heat diffusion transcends boundaries, which can be seen in a simple experiment in which two bodies of the same mass elastically interact with each other:

	Body 1	Body 2
Impulses before collision:	→	←
Impulses after collision:	←	→

Table 3.1: Elastic collision of two bodies

The bodies indicated by arrows in table 3.1 limit each other in the continuation of their local state by the repulsive force existing between them, but not in the continuation of their state of motion. Rather, their respective momentum continues unhindered in the respective other body after the impact in space. And one can even go one step further and put the surface radiation of solid bodies into the context of the trans-boundary momentum continuation, because the emitted radiation takes over the outward directed momentum of the peripheral particles and carries it also beyond the boundary of the periphery of the solid body, which leads to the theory that also, and especially, solid bodies have to radiate for the sake of their stability, because otherwise peripheral particles would detach with higher probability and as a consequence weaken the stability of the solid body.

3.3 Namespace complexity

3.3.1 Finite ordered name spaces

The goal of this section is to deepen the principle of abstract differentiation established in the last section, which is shown in Boltzmann's formula 1.2 for entropy. Let therefore be given an (abstract) set with N pairwise distinguishable elements. Each element is uniquely represented by a name, i.e. the name assignment defines a bijective mapping of the set to the set of names. The only requirement for the names is that they are pairwise distinguishable just like the elements themselves. The mapping, but also the set of names itself, is to be called namespace. The question arises for the most effective namespaces with the shortest names, and it is immediately postulated that number systems form models of effective namespaces. Without justification it is further assumed that for large namespaces the efficiency is independent of the base b of the number system.

The name assignment therefore consists in the fact that the elements of the set are numbered, whereby at the same time the order of the natural numbers is put on the set. The length of the name, which is also called name complexity, is equivalent to the number of digits without consideration of leading zeros. By looking at number systems it becomes clear that namespaces always have short and long names. To examine the total length L_b of such a namespace with $N = b^n$ names, i.e. the sum of all name lengths, it can be noted that for a given name length i there are exactly

$$(b-1)b^{i-1} \tag{3.1}$$

numbers $\neq 0$ with this name length, namely $(b-1)$ digits $\neq 0$ at position i and in each case exactly b digits at the remaining $(i-1)$ positions, so that for the sum of all name lengths and thus the complexity of the whole name space applies:

$$\begin{aligned}
L_b &= \sum_{1 \leq i \leq n} i(b-1)b^{i-1} \\
&= (b-1) \sum_{1 \leq j \leq n} \sum_{j \leq i \leq n} b^{i-1} \\
&= (b-1) \sum_{1 \leq j \leq n} \left[\sum_{j \leq i \leq n} b^{i-1} - \sum_{1 \leq k \leq j-1} b^{k-1} \right] \\
&= (b-1) \sum_{1 \leq j \leq n} \left[\frac{1}{b-1}(b^n - 1) - \frac{1}{b-1}(b^{j-1} - 1) \right] \\
&= \sum_{1 \leq j \leq n} (b^n - b^{j-1}) \\
&= nb^n - \frac{1}{b-1}(b^n - 1)
\end{aligned}$$

With $n = \log_b(N)$ we get[3]

$$L_b = N \log_b(N) - \frac{1}{b-1}(N-1) \tag{3.2}$$

and, if N is sufficiently large, for the mean name length:

$$\frac{L_b}{N} \approx \log_b(N) \tag{3.3}$$

Thus, the averaged name length is exactly the same as the length of the longest name, which illustrates the small proportion of short names that omit leading zeros in a large namespace.

In summary, the complexity of a large namespace of N names is equal to $N \log(N)$ and the average complexity is equal to $\log(N)$. Within such a namespace – with fixed size – the distinction between short and long names is meaningless with respect to the overall complexity.

---

[3] Cf. the following result for L_b with the later defined entropy function $x \mapsto x\log(x) - (x-1)$, shown in figure 7.1

Furthermore, in the number system of base b for number m containing i digits

$$b^{i-1} \leq m < b^i$$

holds and thus also:

$$i - 1 \leq \log_b(m) < i$$

Because even $i - 1 < \log_b(m)$ holds for almost all numbers m with i digits, we get:

$$\sum_{1 \leq i \leq n} (i-1)(b-1)b^{i-1} < \sum_{m=1}^{b^n - 1} \log_b(m) < \sum_{1 \leq i \leq n} i(b-1)b^{i-1} = L_b$$

Because for the expression on the left side

$$\sum_{1 \leq i \leq n} (i-1)(b-1)b^{i-1} = L_b - (b-1) \sum_{1 \leq i \leq n} b^{i-1} = L_b + 1 - b^n = 1 + N(\log_b(N) - 1) - \frac{1}{b-1}(N-1)$$

applies, we arrive at

$$1 + N[\log_b(N) - 1] - \frac{1}{b-1}(N-1) < \sum_{m=1}^{b^n - 1} \log_b(m) < N \log_b(m) < N \log_b(N) - \frac{1}{b-1}(N-1)$$

and thus for large N

$$\log_b(N!) = \sum_{m=1}^{N} \log_b(m) \approx N \log_b(N) \tag{3.4}$$

which also results from the much more accurate Stirling formula:

$$N! = \sqrt{2\pi} N^{N + \frac{1}{2}} e^{-n}$$

3.3.2 Structured Finite Namespaces and the Shannon Entropy

The finite set $\mathcal{M}$, which is represented by a namespace and contains N elements, may now be decomposed into subsets (equivalence classes) with n_1,n_2,n_3... elements, such that

$$\sum_i n_i = N$$

holds. The subsets may each be represented by their own (local) namespaces with complexities $n_i \log(n_i)$ and the original global namespace may be replaced by a structured namespace based on the local names. The global name of an element is then composed of, for example, a global prefix of length $[\log(N) - \log(n_i)]$ (as the name for the respective subset) and the local name of length $\log(n_i)$ for the distinction within the subset. Thus, the total name length $[\log(N) - \log(n_i)] + \log(n_i) = \log(N)$ of an element remains unchanged, and the global complexity of all elements of a subset is:

$$n_i[\log(N) - \log(n_i)] + n_i \log(n_i)$$

Note: The subset name with representation cost $\log(N) - \log(n_i) = \log(N/n_i)$ is short for large subsets and long for small subsets, corresponding to a temporal interpretation according to which division into subsets does not affect the efficiency of global names, i.e., frequently used prefixes are short, rarely used ones are long.

If we sum up the global components of the names, the prefixes of length $[\log(N) - \log(n_i)]$, over all elements of the set $\mathcal{M}$, we get

$$\sum_i n_i[\log(N) - \log(n_i)] = -N \sum_i \frac{n_i}{N} \log\left(\frac{n_i}{N}\right) = -N \sum_i p_i \log(p_i) \tag{3.5}$$

with the relative sizes or probabilities:

$$p_i = \frac{n_i}{N} \tag{3.6}$$

By construction, therefore, vice versa also

$$-N \sum_i p_i \log(p_i) = N \log(N) - \sum_i n_i \log(n_i) \tag{3.7}$$

holds along with the following interpretation:

Global complexity: From the global perspective, there is first a globally valid namespace with N names, corresponding to the N elements of the set $\mathcal{M}$ to be distinguished. According to Eqn. 3.3, the complexity of the namespace is equal to $N \log(N)$. If the set $\mathcal{M}$ is decomposed into classes, each with n_i elements, their own local namespaces, and local complexities $n_i \log(n_i)$, then $-N \sum_i p_i \log(p_i)$ is the global complexity for globally distinguishing the local namespaces.

Globalization complexity: From a local perspective, a namespace with n names has complexity $n \log(n)$. If, from a global perspective, there are a number of such namespaces, each with n_i own names and complexities $n_i \log(n_i)$, then a common global namespace can be formed based on these local namespaces for all $\sum n_i$ elements at once. Merging the local namespaces into the global namespace requires the additional complexity $-N \sum_i p_i \log(p_i)$, which cannot be assigned to one or part of the local namespaces, but only to all of them in common, and which is therefore the global complexity of merging the local namespaces into a global namespace and thus the effort of globalization.

Finally, the globalization complexity $-N \sum_i p_i \log(p_i)$, after division by N, yields the average globalization complexity, which is consistent with Shannon entropy:

$$-\sum_i p_i \log(p_i) \tag{3.8}$$

(Cf. *Gibbs entropy*[4] and *entropie in information theorie*[5])

For further interpretation of the abstract expression 3.8 it is essential to note how increase and decrease of its value can be completely analyzed already from the mutual magnitude relation of only two probabilities p_j und p_k, because as long as the sum $p_j + p_k$ of the two probabilities is constant, also all other probabilities can remain unchanged. I.e. the total complexity breaks down into two parts

---

[4] https://en.wikipedia.org/wiki/Entropy_(statistical_thermodynamics)#Gibbs_entropy_formula

[5] https://en.wikipedia.org/wiki/Entropy_(information_theory)

$$-\sum_i p_i \log(p_i) = -\sum_{i \neq j,k} p_i \log(p_i) + [-p_j \log(p_j) - p_k \log(p_k)]$$

with the constant residual $-\sum_{i \neq j,k} p_i \log(p_i)$ and the variable part $-p_j \log(p_j) - p_k \log(p_k)$ subject only to the constraint $p_j + p_k = const.$ Thus, let $p_i = p + \xi$ and $p_k = p - \xi$ with constant p and variable ξ, then the summed complexity part of these two probabilities is equal to

$$-(p-\xi)\log(p-\xi) - (p+\xi)\log(p+\xi) \tag{3.9}$$

The derivative of this function as a function of ξ gives the expression

$$\log\left(\frac{p-\xi}{p+\xi}\right) = \log(p_k) - \log(p_i)$$

which is positive for $-p < \xi < 0$ and negative for $0 < \xi < p$, so that 3.9 rises and falls monotonically, respectively, and thus reaches its maximum at $\xi = 0$ and its minimum at $\xi = \pm p$ (with the continuous closure $x\log(x) \to 0$ for $x \to 0$)). For the two probabilities p_i and p_k this means that the sum of their complexity components increases as long as they move towards each other, that it is maximal when both have the same magnitude and minimal when one of the two probabilities vanishes in favor of the other.

The following graphs show the Shannon entropy according to 3.9 for the natural logarithm:

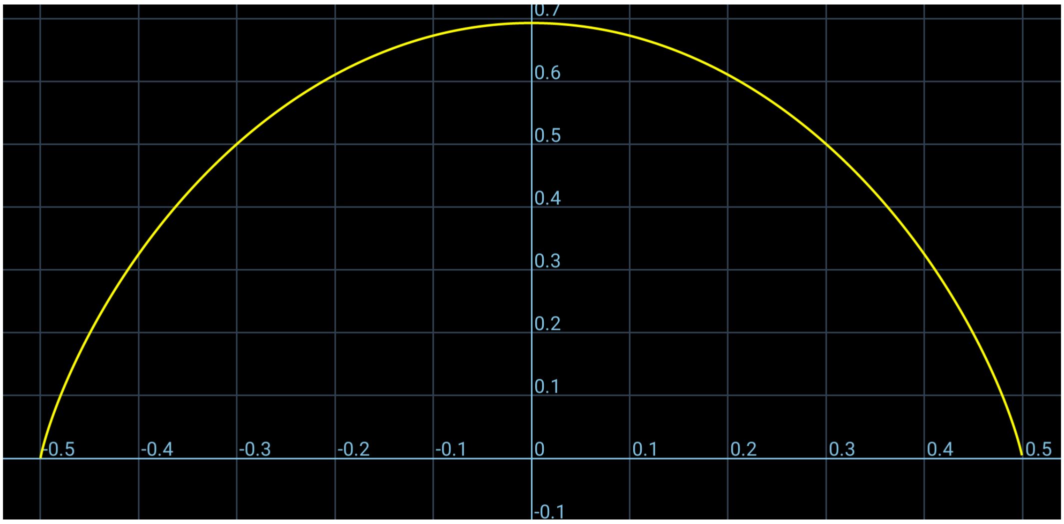


Figure 3.1: Shannon entropy according to Eqn. 3.9 for $p = 0.5$

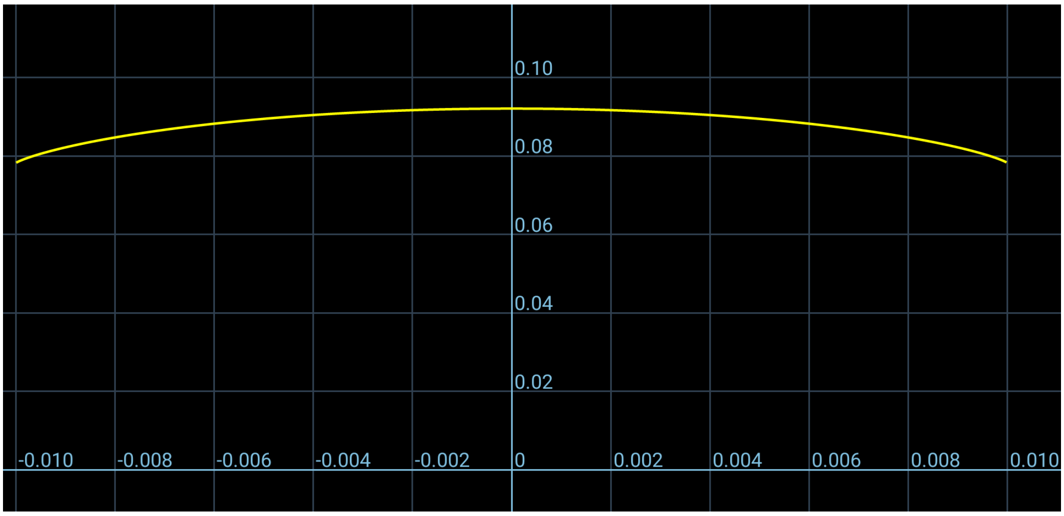


Figure 3.2: Shannon entropy according to Eqn. 3.9 for $p = 0.01$

It also follows from what has been said that for a fixed number of equivalence classes, i.e., a given number n of probabilities p_i, the complexity is maximum if all probabilities have the same size $1/n$, so that for the maximum at n probabilities we get:

$$-\sum_{i=1}^{n} p_i \log(p_i) \leq -\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) = \log(n) \tag{3.10}$$

The same result is given by Jensen's inequality 3.30. From it follows, because the logarithm is a concave function:

$$-\sum_{i=1}^{n} p_i \log(p_i) = \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right) \leq \log\left(\sum_{i=1}^{n} p_i \frac{1}{p_i}\right) = \log(n)$$

In this inequality, according to section 3.4.2, equality holds exactly when all p_i are equal, i.e., exactly in the case of uniform distribution.

By construction in Eq. 3.5, Shannon entropy is non-negative. This also follows from Jensen's inequality, this time because of the convexity of the negative logarithm:

$$\sum_i p_i(-\log(p_i)) \geq -\log\left(\sum_i p_i^2\right) \geq -\log\left(\sum_i p_i\right) = 0 \tag{3.11}$$

In addition to the interpretation given in 3.3.2 as globalization complexity, Shannon complexity can also be interpreted in the following ways:

Representation effort: Since $-\log(p_i) = \log(N/n_i)$ is the representational effort of the i.th equivalence class and therefore $p_i(-\log(p_i))$ is the weighted effort, $-\sum_i p_i \log(p_i)$ is the weighted cumulative effort.

Measure of dissolution and structural decay: When an equivalence class decomposes into two classes so that the elements of each become distinguishable, Shannon entropy grows.

Smoothing measure: The total share of two equivalence classes in the Shannon entropy grows when their sizes converge until it reaches its maximum in the case of equality of both.

Especially in the case of a fixed number of equivalence classes it can be interpreted as a smoothing measure. It measures the extent of equal distribution.

Propagation or size measure: If an initially ignored equivalence class is to be considered after all, i.e. its probability grows from zero to a value greater than zero, then this can only be at the expense of the probabilities of other classes with the consequence that the Shannon entropy increases. Such a process can be interpreted both in the sense of a spreading and in the sense of a growth, the process of smoothing finally as its gradual completion. In fact, dissolution, propagation and growth processes are not distinguished from each other by Shannon entropy. Unlike smoothing, in all three cases the respective process is reflected in the enlargement of the probability distribution.

Temporal complexity: If the elements of the set are not interpreted spatially but temporally or spatio-temporally, namely as possible events of an event space e.g. in the form of states of a state space with probabilities of occurrence p_i, then the Shannon entropy can be understood as that amount of information for a system which would be needed on average to predict the respective next event of the system or e.g. to know the current state of the system. If the information is not available, its quantity is a measure of the lack of knowledge about the current state.

Feature diversity: If one considers the elements of an equivalence class as bearers of a class-specific feature (which is the basis of the division into equivalence classes in the first place, if the equivalence relation is defined via the features, i.e. two elements are equivalent, indistinguishable or without distinctive relevance exactly if they share the same feature), then the complexity is a measure for the feature diversity of the system.

Finally, another significant property of Shannon entropy becomes apparent:

Equality of intensive and extensive complexity: Shannon entropy cannot distinguish between extensive expansion and intensive dissolution processes, because it only has the abstract probabilities. I.e. it increases equally and even indiscriminately by expansion outward as by dissolution inward. Therefore, it is not an extensive measure of size, as section 3.5.2 shows.

3.3.3 Complexity in partition hierarchies and locality

Generalization of Shannon entropy

Each of the equivalence classes of the previous section can in turn be decomposed into equivalence classes, just as the set $\mathcal{M}$ can. Let p_i be first the relative size of the equivalence class $\mathcal{M}_i \subset \mathcal{M}$ and then let p_{ij} with $\sum_j p_{ij} = 1$ be the relative sizes of their own equivalence classes $\mathcal{M}_{ij} \subset \mathcal{M}_i$. These are also equivalence classes in $\mathcal{M}$ with relative magnitudes $p_i p_{ij}$, and relative to $\mathcal{M}$ their cumulative contribution to entropy is equal to $-\sum_j p_i p_{ij} \log(p_i p_{ij})$, for which, due to the additivity $\log(p_i p_{ij}) = \log(p_i) + \log(p_{ij})$ of the logarithm, the following *global-local* relation holds:

$$-p_i \log p_i + p_i \left(-\sum_j p_{ij} \log(p_{ij}) \right) = -\sum_j p_i p_{ij} \log(p_i p_{ij})$$

The cumulative part $-\sum_j p_i p_{ij} \log(p_i p_{ij})$ of all subsets $\mathcal{M}_{ij}$ of $\mathcal{M}_i$ in relation to $\mathcal{M}$ is thus identical to a sum, which consists of the entropy part $-p_i \log p_i$ of the undivided set $\mathcal{M}_i$ in relation to $\mathcal{M}$ and the p_i weighted entropy of the set $\mathcal{M}_i$ itself, which is a local entropy.

If all equivalence classes $\mathcal{M}_i$ of $\mathcal{M}$ are decomposed in this way, then for the global, i.e., total entropy S holds:

$$S = \sum_{ij} p_i p_{ij} \log(p_i p_{ij}) = -\sum_i p_i \log p_i + \sum_i p_i \left(-\sum_j p_{ij} \log(p_{ij}) \right) \tag{3.12}$$

The described two-level decomposition into equivalence classes is a decomposition hierarchy, which we also call division or partitioning hierarchy, which already starts with $\mathcal{M}$ itself, which is the root of the hierarchy, which also forms the 0-th level of the hierarchy or, in another word, its 0-th scale. The equivalence classes $\mathcal{M}_i$ of the first division together form the first scale and the classes $\mathcal{M}_{ij}$ of the second division form the second scale. Now there are in principle no limits to the further recursive continuation of the division. The global-local relationship of Eqn. 3.12 is then – via mathematical induction – extended to the following relationship that is valid for arbitrary partition hierarchies:

$$S = \sum_i p_i \left(-\log(p_i) + \sum_j p_{ij} \left(-\log(p_{ij}) + \sum_k p_{ijk} \left(-\log(p_{ijk}) + \dots \right) \right) \right) \tag{3.13}$$

$$= -\sum_i p_i \log(p_i) - \sum_{ij} p_i p_{ij} \log(p_{ij}) - \sum_{ijk} p_i p_{ij} p_{ijk} \log(p_{ijk}) - \dots \tag{3.14}$$

$$= -\sum_{ijk\cdots} p_i p_{ij} p_{ijk} \cdots \log(p_i p_{ij} p_{ijk} \cdots) \tag{3.15}$$

What we have called above the *global-local* relationship is the reflection of a connection among the formed equivalence classes, which is reflected in general form in the equations 3.13 and 3.15. The connection is that each class, e.g., the class $\mathcal{M}_{ijk}$, is not merely equivalence class of its immediate parent class $\mathcal{M}_{ij}$, from which it emerges by division. Rather, the set of all equivalence classes defined at each scale, initially only locally, always also forms a global decomposition of $\mathcal{M}$. I.e. for the classes $\mathcal{M}_{ijk}$, e.g., of the third scale, by definition initially only holds:

$$\forall ij : (k \neq k' \Rightarrow \mathcal{M}_{ijk} \cap \mathcal{M}_{ijk'} = \varnothing) \wedge \bigcup_k \mathcal{M}_{ijk} = \mathcal{M}_{ij}$$

However, it is true – without proof – as well:

$$((i,j,k) \neq (i',j',k') \Rightarrow \mathcal{M}_{ijk} \cap \mathcal{M}_{i'j'k'} = \varnothing) \wedge \bigcup_{ijk} \mathcal{M}_{ijk} = \mathcal{M}$$

Thus, the deep partitioning hierarchy always defines also a hierarchy with only two scales, namely the 0-th scale consisting only of the set $\mathcal{M}$, and a first scale which I call the (implicit) flat decomposition. Now this relation is reflected in that of the two equations 3.13 and 3.15 for the Shannon entropy. This gives rise to the following possibility:

For a partitioning hierarchy with possibly infinitely many scales, the incremental characteristic of the Shannon entropy as a function of scale can also be studied on the simplified basis of a hierarchy of only two scales according to Eq. 3.12. I.e., actually the Shannon entropy of each two scales n and $n+1$ would need to be calculated according to Eq. 3.13 to finally find the difference of the two results. Instead, one can also resolve the hierarchy up to and including scale n into its flat partitioning and then consider the change in entropy that results from dividing the equivalence classes of this flat partitioning. The following sections give examples of this.

Scale-dependent growth of Shannon entropy

Let $S(n)$ be the Shannon entropy for a partitioning hierarchy as a function of scale n. Then holds:

$$\Delta S(n) := S(n) - S(n-1) \geq 0 \tag{3.16}$$

The Shannon entropy thus grows monotonically as a function of scale. The proof of this follows directly from Eq. 3.12, because $S(n) = -\sum_i p_i \log p_i$ is the entropy of the implicit flat partitioning, and therefore

$$\Delta S(n) = S(n) - S(n-1) = \sum_i p_i \left(-\sum_j p_{ij} \log(p_{ij}) \right) \tag{3.17}$$

holds such that $\Delta S(n)$ equals an average of local non-negative Shannon entropies and is therefore itself non-negative.

Thus, one can also already say that the entropy increase $\Delta S(n)$ at each scale is of the same order of magnitude as the division entropies $-\sum_j p_{ij} \log(p_{ij})$ of the divided equivalence classes $\mathcal{M}_i$.

Infinite growth with bi-partitioning

The point of this section is to determine, since Shannon entropy grows monotonically, whether its growth can be bounded and under what circumstances. From Eqn. 3.10 it follows that it grows logarithmically as a function of the number of equivalence classes in the case of uniform distribution and thus – in this case – also infinitely. In this section we only consider recursively bi-partitioned sets, because such bi-partitioning is characterized by particular simplicity and symmetry. Let, then, be given a binary partitioning hierarchy in which the recursive division of equivalence classes divides each in turn into exactly two classes, so that the hierarchy corresponds to a binary tree. The above equation then yields the following equation for $\Delta S(n)$:

$$\Delta S(n) = \sum_i p_i(-p_{i1} \log(p_{i1}) - p_{i2} \log(p_{i2})) \tag{3.18}$$

So we investigate – as already following the definition 3.8 of Shannon entropy – the entropy change on the basis of bipartitions. Now, in principle, we can already state that the restriction to bipartitions does not imply any loss of generality, because any other – at least finite – division can be obtained by bipartitions carried out in succession.

Because $\Delta S(n)$ according to Eqn. 3.18 is the simple average of the local two-part entropies $\Delta S_i = -p_{i1} \log(p_{i1}) - p_{i2} \log(p_{i2})$, it suffices to study these entropies, which we can now represent – as in the expression 3.9 – because of $p_{i1} + p_{i2} = 1$ with the fixed value $p = 1/2$ as follows:

$$\begin{aligned}
\Delta S_i &= -p_{i1} \log(p_{i1}) - p_{i2} \log(p_{i2}) && (3.19)\\
&= -\left(\frac{1}{2} - \xi\right) \log\left(\frac{1}{2} - \xi\right) - \left(\frac{1}{2} + \xi\right) \log\left(\frac{1}{2} + \xi\right) && (3.20)\\
&= \log(2) - \frac{1}{2}((1-2\xi)\log(1-2\xi) + (1+2\xi)\log(1+2\xi)) && (3.21)
\end{aligned}$$

The function of the expression in Eqn. 3.20 as a function of x is already represented by figure 3.1 – and again by the following figure, this time together with its derivative

$$\log\left(\frac{1-2\xi}{1+2\xi}\right) \tag{3.22}$$

in green color:

Figure 3.3: Entropy increment ΔS_i (yellow) together with its derivative (green) according to 3.22

The maximum of $\Delta S_i = \log(2)$ is reached at $\xi = 0$, which corresponds to the case of equal sizes $p_{i1} = p_{i2} = 1/2$ and, insofar as true for all divisions on all scales, to the uniform distribution, which we already know leads to infinite growth. Here we can see this from the fact that according to Eqn. 3.18 then also for the total entropy growth at each scale

$$\Delta S(n) = \sum_i p_i \log(2) = \log(2)$$

applies, so that the entropy summed over the scales

$$S(n) = \sum_{i=1}^{n} \Delta S(i) = n \log(2)$$

grows linearly as a function of scale n (which is equivalent to logarithmic growth under uniform distribution according to Eqn. 3.10). So

$$\lim_{n\to\infty} S(n) = \log(2) \lim_{n\to\infty} n = \infty$$

holds and hence the growth of entropy is unbounded.

Now one can see from the yellow drawn graph for ΔS_i in the figure above that nothing changes significantly in the growth if one gives up the strict uniform distribution in favor of possible unequal distributions $\xi \neq 0$, as long as these do not really approach the perfect inequality distribution with $\xi = 1/2$ and $\xi = -1/2$, because then ΔS_i and thus also the averaged entropy growth $\Delta S(n)$ for each scale is indeed smaller than $\log(2)$. But as long as these have downward bounded values, the entropy $S(n)$ nevertheless grows infinitely with increasing scale n. Now obviously exactly this is difficult, to get arbitrarily close to the abscissa on the yellow graph above and at the same time to keep sufficient distance from the edges $\xi = \pm 1/2$, which is further complicated by the diverging derivative at these two locations.

Finally, an exact estimate is needed so that with a small $\epsilon > 0$ we set a lower and upper bound

$$-\frac{1}{2} + \epsilon < \xi < \frac{1}{2} - \epsilon$$

for the inequality distribution measure ξ, which is equivalent to:

$$\frac{1}{2} + \xi > \epsilon \wedge \frac{1}{2} - \xi > \epsilon$$

Since we assume that ϵ is in a small neighborhood of 0, and the function $\xi \to -\xi \log(\xi)$ is strictly monotonically increasing there, as is also shown in the figure 3.4 below, it follows from Eq. 3.20 together with the last two relations

$$\Delta S_i \geq \Delta S_{min} := -2\epsilon \log(\epsilon) \tag{3.23}$$

and thus, as already above for the uniform distribution, first $\Delta S(n) \geq -2\epsilon \log(\epsilon)$ and then

$$S(n) \geq -2n\epsilon \log(\epsilon) \tag{3.24}$$

and from it finally:

$$S = \lim_{n\to\infty} S(n) \geq \lim_{n\to\infty} -2\epsilon \log(\epsilon) n = -2\epsilon \log(\epsilon) \cdot \lim_{n\to\infty} n = \infty \tag{3.25}$$

Thus, it is clear that entropy grows infinitely with scale when $\epsilon > 0$ is a fixed finite value. To proceed, let us first specify the objective. It consists in,

- to exclude the perfect relative non-uniform distribution with which one of the quantities p_{i1} or p_{i2} equals zero and the other equals 1 and thus the entropy increase $\Delta S_i = 0$ vanishes, and
- nevertheless limit the entropy growth as a function of scale.

The perfect non-uniform distribution is nothing else than an improper division of the set $\mathcal{M}_i$ into the empty set and the improper subset $\mathcal{M}_i \subseteq \mathcal{M}_i$. Now, if the goal is really to exclude only this case, then, however, any fixed value $\epsilon > 0$ is more than actually necessary for achieving this goal, because any smaller value corresponds better to the goal. Therefore, one can try to replace the fixed value by an infinitesimal defined by a weakly decreasing zero sequence. Accordingly let

$$\epsilon := \epsilon_1, \epsilon_2, \epsilon_3 \cdots \tag{3.26}$$

be a strictly monotonous but weakly decreasing zero sequence. More precisely:

$$(\forall n \in \mathbb{N} : \epsilon_{n+1} < \epsilon_n) \wedge \left(\lim_{n\to\infty} \epsilon_n = 0\right) \wedge \left(\lim_{n\to\infty} \epsilon_n \cdot n = \infty\right) \tag{3.27}$$

In this definition, we connect n to the scale of the partitioning hierarchy in the sense that the limits $\lim_{n\to\infty} S(n)$ and $\lim_{n\to\infty} \epsilon_n$ are taken together. The terms ϵ_n, like ϵ and ξ, have the meaning of a relative share of an equivalence class, which in the case of the infinitesimal we expect to represent a share smaller than any $\epsilon > 0$. On the other hand, we need a criterion to sufficiently delimit its value from 0, since we want to exclude $\epsilon = 0$ and the associated improper division according to the above. For this we choose the scale limit $\lim_{n\to\infty} 1/n = 0$, which defines the value of 0 in the continuum to first order. With respect to this definition, the third condition $\lim_{n\to\infty} \epsilon_n \cdot n = \infty$ in 3.27 then has the meaning of said demarcation, which is valid on all scales.

Accordingly, 3.25 becomes:

$$S = \lim_{n\to\infty} S(n) \geq 2 \lim_{n\to\infty} n\epsilon_n \cdot (-\log(\epsilon_n)) = \infty \tag{3.28}$$

The limit is infinite, since the values $n\epsilon_n$ and $-\log(\epsilon_n)$ each grow infinitely according to the definition of the infinitesimal.

In summary, we can conclude that the Shannon entropy for an infinite binary partitioning hierarchy is infinite even when partitioning into subsets of even infinitesimal relative size only is allowed.

Infinite growth in the general case

For the general case, we again take from Eqn. 3.17 that for the entropy increase $\Delta S(n)$ on a scale

$$\Delta S(n) = \sum_i p_i \Delta S_i$$

with $\Delta S_i = -\sum_j p_{ij} \log(p_{ij})$ holds. Thus, the summands of ΔS_i all have the form of the entropy function $\xi \mapsto -\xi \log(\xi)$, which the following figure shows together with its derivative.

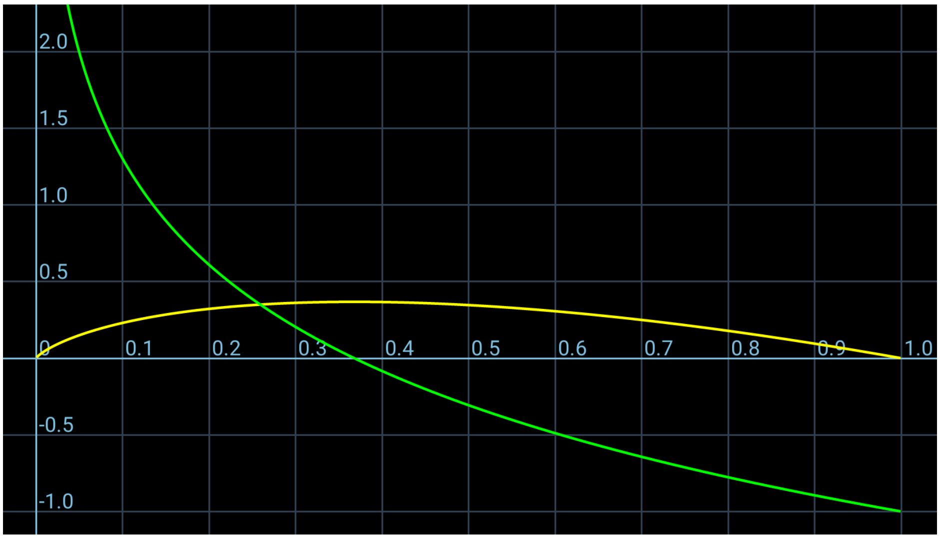


Figure 3.4: Graph of the function $\xi \mapsto -\xi \log(\xi)$ (yellow) with its derivative (green)

At least one of the summands has the smallest size $p_{ij} \leq 1/2$, so that its entropy $-p_{ij}\log(p_{ij})$ is determined by the function values in the left part of the domain of the entropy function. Therefore, there is also at least one summand whose magnitude – except for the factor 2 – is subject to the same estimation as that for ΔS_i of the two-part system according to Eqn. 3.23. Since this summand is the smallest, it follows

$$\Delta S_i \geq -\epsilon \log(\epsilon)$$

again with a lower bound ϵ for the quantities p_{ij} to exclude the case $\mathcal{M}_{ij} = \varnothing$. Everything else follows as in the estimation for the two-part system, so that Shannon entropy is not bounded also in the case of general hierarchy partitioning, even if *almost* improper partitions based on an infinitesimal are permitted.

3.3.4 Complexity as a cost function and gradual distinctiveness

Difference, ratio, relative and logarithmic difference

For two natural numbers m and n the following holds:

$$m = n \Leftrightarrow m - n = 0$$

The arithmetic difference $m - n$ thus allows to check the equality of two natural numbers. However, it goes beyond the mere identity check and even allows the determination of a gradual difference. This difference is independent of the absolute magnitudes of the numbers, independent therefore of where in the ordered set of the Natural Numbers both have their place in relation to the reference point, the zero, or some other third point. The most important and at the same time most elementary difference is that between a number n and its predecessor $n-1$ or vice versa between n and its successor $m := n+1$. Thus the difference allows the test of the successor relation of the natural numbers:

$$m = n + 1 \Leftrightarrow m - n = 1$$

In a similar way – as above for the difference – holds for the ratio if $n \neq 0$:

$$m = n \Leftrightarrow \frac{m}{n} = 1$$

So also the relation m/n allows the test of equality and likewise goes beyond the mere identity test by allowing the determination of a gradual difference. In contrast to the difference, however, in the ratio the absolute quantity, i.e. their respective relation to the reference point, the zero, can be mirrored, and increasingly so, if their difference is small, but not equal to zero, which in turn applies in particular to a number n and its successor $n+1$:

$$\frac{n+1}{n} = 1 + \frac{1}{n}$$

The difference of two neighbors measured by the ratio is thus the smaller, the larger they are, i.e. the larger their difference to zero is. In general, if the difference between m and n is not large in relation to n in any case, the ratio m/n actually expresses a relation between the three numbers $0, n, m$.

Third, the relative difference connects the difference $m - n$ with the ratio m/n:

$$\frac{m-n}{n} = \frac{m}{n} - 1$$

For it, instead of the above equivalence, holds:

$$m = n \Leftrightarrow \frac{m}{n} - 1$$

The logarithmic difference connects the ratio with the difference in a special way:

$$\log(m) - \log(n) = \log\left(\frac{m}{n}\right)$$

For large numbers m, n with relatively small difference $m - n$, so that $x := m/n \approx 1$, the relative and the logarithmic difference have similar characteristics, which can be seen in the graphs of the two functions $x \mapsto \log(x)$ (yellow) and $x \mapsto x - 1$ (green) in the neighborhood of $x = 1$:

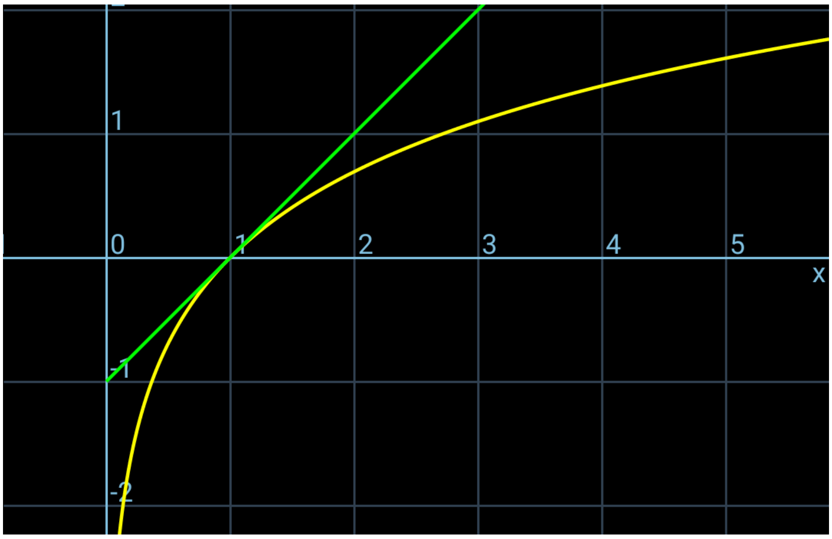


Figure 3.5: Verwandtschaft der relativen und logarithmischen Differenz

Both differences can be understood as a gradual measure of the distinguishability of two numbers m and n, or of the distinguishability of the finite sets $\{0, 1, 2, ...m\}$ and $\{0, 1, 2, ...n\}$ represented by the numbers m and n. Both differences can take positive as well as negative values, negative values if the ratio of the numbers is reversed and $m/n > 1$ becomes $m/n < 1$. However, it is precisely in this case that the logarithmic difference is distinguished from the relative difference by symmetry, which is more precisely an anti-symmetry. Because, if the numbers are exchanged as in $m \to m^{'} = n$ and $n \to n^{'} = m$, we get

$$\frac{m}{n} - 1 \to \frac{m^{'}}{n^{'}} = \frac{n}{m} - 1$$

but:

$$\log(m) - \log(n) \to \log(m^{'}) - \log(n^{'}) = \log(n) - \log(m) = -(\log(m) - \log(n))$$

So in both cases there is a sign change. The logarithmic difference, however, performs the exchange completely symmetrically beyond this, as is also shown by the two graphs in Figure 3.5, because the relative ratio is constrained below for ratios $x \leq 1$, namely by -1, but not for ratios $x > 1$. The logarithmic difference, on the other hand, is equally unbounded in both cases.

Complexity and gradual distinctiveness

The meaning of gradual distinguishability is revealed by the count

$$1, 2, 3 \cdots \mathit{Many}$$

which is real and lived everyday life. *Many* is the vestibule of infinity, whose striking feature is perspectival indistinguishability or the – perhaps supposed – lack of need for distinction.

At the end of the section 3.3.1 on finite ordered namespaces, the observation was made that for sets of fixed size N the distinction between short and long names is meaningless and instead all names can be assigned the same, average complexity $\log(N)$. For infinitely growing sets or infinite sets like the natural and integer numbers, however, the exact assignment of complexity to subsets or the elements themselves is mandatory:

More precisely, in the sense of section 3.3.1, complexity is the consequence of the simple binary pairwise distinction (yes or no) of elements of a finite set based on a number system.

Now also the natural numbers themselves arise on the basis of simple binary distinction,[6] which are characterized by a global asymmetry, namely the asymmetry of order and for instance the asymmetry between finite and infinite part.[7] This asymmetry is also revealed in its representation by number systems in the form of complexity. According to it, each section $0, 1, 2 \cdots n$ of the natural numbers has the (average) complexity $\log(n)$, so that sections of different length are evaluated differently by complexity. It is essential that the complexity increase is perspectival and takes place in relation to a reference point, zero, but that it is compelling if the perspective (reference point) remains the same and that every change of perspective does not change anything globally.

Obviously, this gradual increase of complexity is connected with the gradual decrease of either the distinguishability or the necessity to distinguish. Thus, the gradual measure of complexity is contrasted with the gradual measure of distinctness or need for distinctness for two numbers m and n. Intuitively, therefore, the most important requirement for such a measure is that for large numbers m, n with small difference $m - n$ it also be small, which is satisfied in the same way by the relative difference as well as the logarithmic difference. However, the asymmetry of the relative difference disqualifies it for a measure of distinctness for arbitrary numbers m and n, while the logarithmic difference combines the properties of the arithmetic difference $m - n$ and the ratio m/n and is probably therefore the ideal measure of distinctness. For ratio values m/n clearly beyond one, the focus of their meaning is rather that of relative complexity, for ratio values in the vicinity of one it is rather the distinctness.

A comparison with Eqn. 3.29 of the following section shows that gradual distinctness is a special case of the complexity $\log(\Delta x/\Delta x_R)$ of the continuum. This too is a relative measure, defining the complexity of Δx relative to the reference Δx_R. And even the originally defined complexity $\log(m)$ is a special case of gradual distinctness with $n = 1$, so any complexity is relative, including the complexity $\log(m)$.

Finally, the necessity of distinction is the other side of gradual distinctness. The use of the indefinite number word *many*, for example, implies that the exact specification of the number is not ascertainable because of its size or – conditioned by the context – not necessary. What may be the case in everyday life with already rather small numbers is in mathematics the definition of the limit

$$\lim_{n\to\infty} n$$

which is reached when continuing implies no difference. The prototype of this limit leads to indistinguishability of adjacent numbers n and $n + 1$ with limit

$$\lim_{n\to\infty} \log\left(\frac{n+1}{n}\right) = 0$$

According to this, the infinite is the space

---

[6] This becomes very clear in the second volume about the continuum and therein about the Natural Numbers and individual distinctness

[7] Here, too, reference can be made to the second volume on the continuum therein to the chapter on extensive infinite space

- in which the distinction of neighboring elements or the size of the sets represented by them is not possible
- or – in the respective context – makes no difference

$\cdots$ and into which, depending on the context, the transition is gradual and to which complexity sets a soft boundary.

Complexity as a cost function

The soft boundary mentioned in the last sentence of the last section suggests that complexity involves a further interpretation. First, complexity is the consequence of the simple binary distinction, as the last section also points out, and leads immediately to the notion of gradual distinctiveness. Complexity itself as well as distinctness both lead to this interpretation, in which complexity is the universal basis of every cost function, which does not set a rigid, but just a soft boundary for the transition to large numbers, which is shown in particular in the decreasing perspective distinctness of the basically binary distinguishable elements of a set, as the natural numbers and the extensive space are.

3.4 Complexity of the continuum

3.4.1 The infinity of the continuum and negative complexity

In a certain way the infinity in the continuum is real in contrast to the extensive infinity of the Natural Numbers, which, one could say e.g. in the shape of the size of the universe is far away and – possibly – has only little importance for the respective local reality. In contrast to this, the intensive infinity of the continuum is local, i.e. not even only in immediate neighborhood, but in the center of all physical being, however large or small this being may be, mathematically expressed for instance by a local environment in space. The complexity of such an environment with all its numbers contained in it is always infinite, even if the environment is restricted to rational numbers. In the case of irrational numbers already the complexity of only a single number is infinite, because its description length is infinite, namely that of an infinite sequence or series.

Now there are obviously two ways to approach the complexity of the continuum. On the one hand, the Shannon entropy offers itself for the time being as a measure of intensive and extensive complexity[8]. For the most essential feature of the continuum is that it sets no limits to dissolution, i.e., to continued division. This essential feature, or rather the measure of this feature, is reflected by Shannon entropy $-\sum_i p_i \log p_i$ in an exemplary way, insofar as, although one starts from a finite namespace of a set $\mathcal{M}$ with finitely many N elements of section 3.3.2, this can be extended indefinitely by the generation of local namespaces during continued division, insofar as division leads to no end. The continuing division process would thus be accompanied by the continuing extension of the namespace, which may be necessary for physical reasons, for example, if a particle cannot be uniquely assigned to a spatial environment (cell), i.e., if the classical infinitesimal mass point must give way to the particle of modern physics with extended propagation in space. Now this characteristic of Shannon entropy is both a curse and a blessing. On the one hand it is the perfect measure of resolution, on the other hand it does not converge in the case of indefinitely continued division, as shown in section 3.3.3.

Now, at the heart of Shannon entropy is the indistinguishability of the elements within each of the equivalence classes. As described, as the equivalence classes become smaller by resolution, the complexity increases because the overall global distinguishability of the elements increases. Let

---

[8] Cf. section 3.5.2 on the intensity of Shannon entropy

us assume that we want to distinguish matter increasingly better on the basis of its assignment to space (cells of space) by decreasing the size of the cells, in order to describe the physical state increasingly better, as it is already described in the introduction. Then, according to what has been said above, in this reduction process the Shannon entropy of matter grows infinitely, if matter can be divided infinitely, as modern physics claims.

On the other hand, however, the complexity of the spatial continuum is to be considered in itself, i.e. independently of its interweaving with matter. The complexity introduced in section 3.3.1 is nothing else than the complexity of the discrete space. In other words, the discrete mathematical-physical space is the natural realization of a namespace as introduced in that chapter, a set namely with a focus on the abstract distinction of its elements. In the following, space itself is considered as a namespace, which thus exclusively serves the purpose of the mere distinction of its elements themselves and everything what stands in a relation with these elements, thus for instance matter. The relation between this namespace and the anthropogenically created namespace of numbers will be treated in the following chapters. Two important results of these chapters are two perspective dependencies, on the one hand the choice of the reference point, zero, and on the other hand the choice of the reference scale or – equivalently – of a unit of length. The one as well as the other choice is on the one hand necessary, on the other hand it breaks the symmetry of space. The importance of the zero shows up in a particularly striking way in the Peano axioms, under which the zero axiom makes the zero an element distinguished from all other elements by the fact that it is the only element which has no predecessor under the successor mapping $n \to n+1$. The resulting symmetry breaking is global, manifesting itself, among other things – from the perspective of each element n – in the distinction of a finite ($\leq n$) and an infinite part ($\geq n$) of the natural numbers. It is due to this perspective distortion that complexity, as defined in section 3.3, also divides space into complex and less complex parts, which gives rise to the phenomenon of complexity in the first place. At the same time, the elements of discrete space thus distinguished are not abstract elements, not even points of space, but extended cells of space, which in turn can be subdivided, that is, resolved. In connection with this is the second symmetry breaking as a result of the determination of a reference scale together with a unit of length and thus the determination of the cell volume. Namely, the symmetry of space is broken by highlighting the cell edges, which are necessary for the definition of the cells, and which are related to the cell volume. The space symmetry is satisfied only by continued division of the cells into smaller rational cells up to real indivisible points of the space. The meaning of the rational and in particular of the irrational numbers is therefore in the end the clearing up of the arbitrary symmetry breaking of the continuous space which is inevitably connected with the natural numbers and every fixed scale.

Let us now assume that, for example, the one-dimensional space, the number line, is decomposed into unit cells of the same size, the reference quantity u. Then neighboring unit cells can in turn be grouped into cell clusters, a neighborhood. Depending on the size of a neighborhood, which is measured by the number n of its cells, it has the local mean complexity $\log(n)$ – based on the fixed size u, according to Eq. 3.3. Now, the possibility of forming such neighborhoods to larger complexes is one side of space, including the discrete space consisting of cells. But the other side of the continuum is now the possibility of division of cells, including unit cells, and the question arises of the complexity of a part of a unit cell. If now the summary of n unit cells has the complexity $\log(n)$ and thus a single cell has the complexity $\log(1) = 0$, then a cell part of the relative size $p \leq 1$ must have the complexity $\log(p) \leq 0$, as far as one wants to ascribe a complexity to a cell part at all. But on the other hand this is necessary, because no cell is basically distinguished from others in the continuum. One could also have declared pu the other way round as the reference quantity. The original unit cell would then be composed of $1/p$ new unit cells and would accordingly have the complexity $\log(1/p) = -\log(p) > 0$. The new unit cell, on the other hand, would have complexity 0.

In the continuum, complexity is therefore necessarily relative and – as has already been mentioned and will show again – not only in the continuum. And therefore the definition of a reference quantity or a reference scale is a prerequisite for the existence of complexity at all. And for the complexity of importance are also only relative quantities like the quantity p used above. I.e. the complexity of any cell of size Δx with respect to the unit cell of size u is:

$$\log\left(\frac{\Delta x}{u}\right) \tag{3.29}$$

One can also interpret the change of complexity from positive to negative values as the change from extensive to intensive complexity of the space, likewise as a change from the interior to the exterior view. Here the perspective is always that of the unit cell. If the unit cell is part of a larger cell, then this is viewed from the inside. If, on the other hand, the unit cell surrounds a smaller cell, the latter is viewed from the outside.

The necessity of the reference to a unit of magnitude has its reason in the scale symmetry of the continuum which distinguishes no scale before other scales. I.e. the continuous space determines in the scale space as little as in itself a reference point and is therefore a space of relative magnitudes only.

For cell sizes Δx converging to zero, i.e., if the limit of infinite resolution is taken, the complexity in Eqn. 3.29 diverges to $-\infty$. This negative divergence is in contrast to the divergence of the Shannon entropy to $+\infty$ at infinite resolution. The sum, however, of both complexities can be bounded, as shown in section 3.4.4 on differential and relative entropy in respect to an upper bound.

3.4.2 Jensen's inequality

For the rest of this work, the following relation, valid for concave functions and known as *Jensen's inequality*[9], will prove essential:

Let f be a real-valued concave function, then for positive real numbers p_i with $\sum_i p_i = 1$ and real numbers x_i in the domain of f the following is true

$$\sum_{i=1}^{N} p_i f(x_i) \leq f\left(\sum_{i=1}^{N} p_i x_i\right) \tag{3.30}$$

with the essential addition that the equality holds exactly if all x_i are equal – independent of the p_i. (For convex functions f the inverse relation $f(\sum p_i x_i) \leq \sum p_i f(x_i)$ holds).

3.4.3 Maximum property of mean values and uniform distribution

When the distribution of p_i is fixed, the function $F : \mathbb{R}^N \to \mathbb{R}$ can be defined as:

$$F(x_1, x_2, \cdots x_N) = \sum_{i=1}^{N} p_i f(x_i)$$

Moreover, if the mean $\sum p_i x_i$ is fixed, then the function F has under this constraint according to Eqn. 3.30 a global maximum at

$$x_1 = x_2 = \cdots x_N = \sum p_i x_i$$

---

[9] https://en.wikipedia.org/wiki/Jensen's_inequality

because if the mean $\sum p_i x_i$ is fixed, the x_i are exactly equal if they are all equal to the mean.

Since the logarithm is concave, according to inequality 3.30 also

$$\sum_{i=1}^{N} p_i \log(x_i) \leq \log\left(\sum_{i=1}^{N} p_i x_i\right) \tag{3.31}$$

is valid, which, as will be shown, implies among other things statements about the complexity of a physical system composed of parts. Here the mean values $\sum p_i x_i$ have the meaning of conservative quantities, particle number and kinetic energy (conservation in the case of the ideal gas), and the x_i have the meaning of particle density and temperature, respectively. The statements are that the complexity of the total system is maximal if the densities or temperatures are equal in all subsystems, i.e. if the particles and the temperatures are equally distributed.

3.4.4 Differential, Relative and Density Entropy

The problem of divergence at unbounded resolution is also immediately apparent when the form of Shannon entropy is transferred to the continuum in terms of a continuous probability density $P(x)$ in an integral. Namely it holds:

$$-\int P(x)\log(P(x)dx)\,dx = -\int P(x)\log(P(x))\,dx - \int P(x)\log(dx)\,dx \tag{3.32}$$

The right side of this equation diverges according to the divergence of $\log(dx) \to -\infty$ for $dx \to 0$. The integral $-\int P(x)\log(P(x))\,dx$ is called *Differential entropy*[10] in contrast to the Shannon entropy $-\sum p_i \log(p_i)$ for finite or at least discrete systems.

Using the *Relative entropy*[11]

$$-\int P(x)\log\left(\frac{P(x)dx}{Q(x)dx}\right)dx = -\int P(x)\log\left(\frac{P(x)}{Q(x)}\right)dx \tag{3.33}$$

$$= -\int P(x)\log(P(x))\,dx + \int P(x)\log(Q(x))\,dx \tag{3.34}$$

based on a second probability density Q it is possible to avoid the divergence. However, instead of the general case of an initially arbitrary complexity $\log(Q(x))$, in the following we will only consider the cell complexity $\log(\Delta x/u)$ according to 3.29. One can weight these complexities in an analogous way as it is done for the relative entropy 3.34 above, with the probabilities p_i of the Shannon entropy $-\sum p_i \log(p_i)$ and finally consider the sum of both complexities, which I call density entropy in the following:

$$S_D := -\sum_i p_i \log(p_i) + \sum_i p_i \log\left(\frac{\Delta x_i}{u}\right) \tag{3.35}$$

For example, in this equation Δx_i is the volume of a cell in position space and p_i is the relative particle share contained therein.

---

[10] https://en.wikipedia.org/wiki/Entropy_(information_theory)#Differential_entropy

[11] https://en.wikipedia.org/wiki/Kullback-Leibler_divergence

It can then be stated that this sum, unlike Shannon entropy alone, is upper bounded – for systems of finite extension – by the system size $L := \sum \Delta x_i$. Namely, it holds because of Jensen's inequality 3.30:

$$S_D = \sum_i p_i \log\left(\frac{\Delta x_i}{p_i u}\right) \leq \log\left(\sum_i \frac{\Delta x_i}{u}\right) = \log\left(\frac{L}{u}\right) \tag{3.36}$$

In this relation, according to section 3.30, the equal sign holds exactly when the ratios $\Delta x_i / p_i$ are all equal, i.e., the probabilities p_i are proportionally equally distributed.

With the values $q_i := \Delta x_i / L$ normalized to 1, the density entropy finally takes the form of the relative entropy, since

$$S_D = -\sum_i p_i \log\left(\frac{p_i}{\frac{\Delta x_i}{u}}\right) = -\sum_i p_i \log\left(\frac{p_i}{q_i}\frac{u}{L}\right) \tag{3.37}$$

it follows:

$$S_D = S_D\left(\frac{L}{u}\right) = \log\left(\frac{L}{u}\right) - \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \tag{3.38}$$

The density entropy is thus the sum of an external (extensive) and an internal (intensive) complexity. The external complexity is determined by the ratio L/u. In a physical interpretation it is the expression of a relation of the system as a whole to the rest of the universe. The sum on the right side in Eqn. 3.38, the actual relative entropy, on the other hand, is the measure of exclusively internal relations represented by the internal relative quantities p_i and q_i and their ratios p_i/q_i. This measure is always ≤ 0 according to the estimation 3.36, but this follows just as well again from Jensen's inequality. If we declare the extensive system size L to be the reference quantity, i.e., set $u = L$, then the complexity of the external relation vanishes, and we obtain the non-positive density entropy of the system normalized to 1:

$$S_D(1) := -\sum_i p_i \log\left(\frac{p_i}{q_i}\right) \leq 0 \tag{3.39}$$

Further, it follows from the left-hand side of Eqn. 3.37 by transition to the integral representation, the relation between the density entropy and the *differential entropy*

$$S_D\left(\frac{L}{u}\right) = -\int_{\dots}^{\dots+L} P(x)dx \log\left(\frac{P(x)dx}{\frac{dx}{u}}\right) = -\log(u) - \int_{\dots}^{\dots+L} P(x)\log(P(x))\,dx \tag{3.40}$$

and with $Q(x) := 1/L$ from the right part of Eqn. 3.37 the relation of S_D with the *relative entropy* according to Eqn. 3.33:

$$S_D\left(\frac{L}{u}\right) = \log\left(\frac{L}{u}\right) - \int_{\dots}^{\dots+L} P(x)\log\left(\frac{P(x)}{Q(x)}\right) dx \tag{3.41}$$

Thus, the density entropy S_D as the sum of the Shannon entropy and the complexity of the continuum is identical to the

- differentiel entropy $-\int P(x)\log(P(x))\,dx$ (with integration over a volume of size L) taking into account the correction constant $-\log(u)$,
- relative entropy, if this calculates the entropy of the distribution P relative to the uniform distribution $Q(x) \equiv L$ of the continuum (also over a volume of size L), taking into account the correction constant $\log(L/u)$, the extensive complexity of the system relative to the reference quantity u.

Since u is the arbitrary fixed value of the length unit of a reference scale, the relation between S_D and the differential entropy depends on this choice, but not the relation between changes of both when changing the probability density, e.g. from P_1 to P_2, in which case even without correction

$$\Delta S_D = -\int P_2(x)\log(P_2(x))\,dx + \int P_1(x)\log(P_1(x))\,dx \tag{3.42}$$

holds.

Finally, Jensen's inequality 3.30 holds analogously for integrals, so that the differential entropy in Eqn. 3.40 is also maximal exactly when all probabilities $P(x)$ have the same magnitude, i.e., when $P \equiv 1/L$ is uniformly distributed. Analogously to Eqn. 3.36 therefore also holds for the differential entropy:

$$S_D\left(\frac{L}{u}\right) \leq -\log(u) - \frac{1}{L}\log(1/L)\int_{\dots}^{\dots+L} dx = \log\left(\frac{L}{u}\right) \tag{3.43}$$

3.5 Extensive and intensive size

3.5.1 Standard deviation

Extensivity in finite systems

Let

$$x : \mathcal{M} \to \mathbb{R}$$

be a function defined on a finite set $\mathcal{M}$. We assume that because of finiteness the elements of $\mathcal{M}$ can be represented by natural numbers $1, 2, \dots N$ and therefore we immediately assume that $\mathcal{M} := 1, 2, \dots N$ is such a finite section of the natural numbers and x is a discrete function

$$x : \mathcal{M} \to \mathbb{R}, i \mapsto x_i$$

on the section. The indexed particles of classical physics, more precisely their indexed location or velocity components along an axis of the position or velocity space, are examples of such a function. For the variance σ_x^2 of the function values, which is defined as the mean of the squared deviations from the function mean, then holds:

$$
\begin{aligned}
\sigma_x^2 &= \frac{1}{N}\sum_i \left(x_i - \frac{1}{N}\sum_j x_j\right)^2 = \frac{1}{N}\sum_i \left(x_i^2 - 2x_i\frac{1}{N}\sum_j x_j + \left(\frac{1}{N}\sum_j x_j\right)^2\right) \\
&= \left(\frac{1}{N}\sum_i x_i^2\right) - 2\left(\frac{1}{N}\sum_j x_j\right)^2 + \frac{1}{N}N\left(\frac{1}{N}\sum_j x_j\right)^2 = \frac{1}{N}\sum_i x_i^2 - \left(\frac{1}{N}\sum_i x_i\right)^2 \\
&= \frac{1}{N}\sum_i \left(x_i^2 - x_i\frac{1}{N}\sum_j x_j\right) = \frac{1}{2N^2}\sum_{ij}(x_i^2 - 2x_ix_j + x_j^2) = \frac{1}{2N^2}\sum_{ij}(x_i - x_j)^2
\end{aligned}
$$

The important relations are:

$$
\sigma_x^2 = \frac{1}{N}\sum_i \left(x_i - \frac{1}{N}\sum_j x_j\right)^2 = \frac{1}{N}\sum_i x_i^2 - \left(\frac{1}{N}\sum_i x_i\right)^2 = \frac{1}{2}\left(\frac{1}{N^2}\sum_{ij}(x_i - x_j)^2\right) \tag{3.44}
$$

Only mean values appear in this equation. In addition to the variance on the left side, in particular, on the right side there is half the mean of the squared differences of all pairs (x_i, x_j) – including such pairs with $i = j$.

Like the position and velocity states, the function x could also be defined with multiple components, such as a function $\mathbf{r} : \mathcal{M} \to \mathbb{R}^3, i \mapsto \mathbf{r}_i = (x_i, y_i, z_i)$, for which – as for the component functions x_i, y_i and z_i – with the scalar product $\mathbf{r}_i \cdot \mathbf{r}_j$ instead of the simple product x_ix_j the relation 3.44 would hold. And it would be $\sigma_{\mathbf{r}}^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2$. For simplicity, however, we consider only one-dimensional systems, because the point is only to make clear the fundamental property of the standard deviation $\sigma_x := \sqrt{\sigma_x^2}$ as an extensive measure of magnitude.

Since there are only finitely many function values, the number of absolute differences $|x_i - x_j|$ is also finite, so there is a pair (i, j) with maximum distance L, which we usually consider as a possible benchmark for the system size. As Eqn. 3.44 shows, the variance does not represent the square of this maximum distance, but is related – according to the right-hand side of the equation – to the mean $M[(\Delta x)^2]$ of the distance squares, given by

$$
M[(\Delta x)^2] := \frac{1}{N(N-1)}\sum_{i\neq j}(x_i - x_j)^2 = \frac{2}{N(N-1)}\sum_{i<j}(x_i - x_j)^2 \tag{3.45}
$$

from which

$$
\sigma_x^2 = \frac{1}{2}\frac{N-1}{N}M[(\Delta x)^2] \tag{3.46}
$$

follows. The correction factors $1/2$ and $(N-1)/N$ have a common reason. Namely, they both reflect the respective priorities of the two definitions. The focus of the variance is on the singular function values x_i with their respective relation to the common mean $1/N\sum_i x_i$, while the mean $M[(\Delta x)^2]$ gives priority to the immediate relations between each two function values in the form of the distance squares and as a consequence priority to pairs of the function values. With this difference in priority the two factors are now related as follows:

- Faktor 1/2: The variance measures the distance squares of the singular function values in relation to the central mean $1/N \sum_i x_i$, while $M[(\Delta x)^2]$ takes into account the distance squares with their full distances.
- Faktor $(N-1)/N$: The right large parenthesis in Eqn. 3.44 shows that, as mentioned above, the variance is also a mean of the distance squares, but it additionally includes the squares $(x_i - x_i)^2$, which, on the other hand, are explicitly excluded in the mean $M[(\Delta x)^2]$. This correction is related to the *Bessel correction*[12] in statistics, which becomes necessary for samples with replacement so as to obtain an expectation-true sample, as shown in *expectation-true sample variance*[13]. Accordingly, sampling experiments to estimate the variance of a random variable differ from those to estimate the variability of the variable based on pairwise differences. In the first case, the performance of each single experiment is the establishment of a single value. In the second case, on the other hand, the single experiment consists in finding two values.

Since the right-hand side of equation 3.44 does not depend on the choice of the reference point, the same is true for the left-hand side, the variance. Therefore, to study the properties of the variance, the reference point can be chosen to coincide with the mean, so that

$$\frac{1}{N}\sum_i x_i = 0 \tag{3.47}$$

applies and thus also:

$$\sigma_x^2 = \frac{1}{N}\sum_i x_i^2 \tag{3.48}$$

This equation states that the variance and with it the standard deviation σ_x disproportionately emphasizes the periphery. That is, it grows monotonically for each i whose value x_i moves away from the reference point, the system center, and approaches the system outer boundary, and therefore reaches its maximum when the system interior is completely emptied in favor of the periphery. With the system outer boundary fixed, represented for example by the limits of an interval $[a, b]$, the standard deviation is therefore at its maximum when the x_i are unevenly distributed with the only two accumulations at a and b, as shown in the following figure:

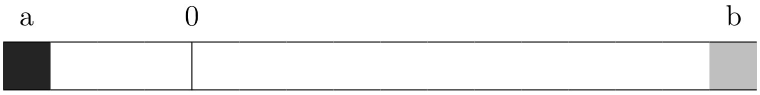


Table 3.2: Periphery-heavy distribution with clusters at system outer boundaries a and b.

Because of Eqn. 3.47, if not all x_i are zero, $a < 0 < b$. Let p and $1-p$ be the relative frequencies for the frequencies of x_i at a and b, respectively. From Eqn. 3.47 then we get $p(-a)+(1-p)b = 0$, giving

$$p = \frac{b}{b-a}$$

Correspondingly, Eqn. 3.48 becomes $\sigma_x^2 = pa^2 + (1-p)b^2$, from which, together with the above equation for p and $L := b - a$

$$\sigma_x^2 = (-a)b = (L-b)b \tag{3.49}$$

---

[12] https://en.wikipedia.org/wiki/Bessel's_correction

[13] https://en.wikipedia.org/wiki/Variance#Unbiased_sample_variance

follows. All variances according to this equation are maximum variances of a respective system configuration given by the system outer boundaries a and b or alternatively by the parameters L and b, among which – for fixed L – again the symmetric configuration with $b = L/2$, shown in table 3.3, has the largest variance $\sigma_x^2 = (1/4)L^2$. For the standard deviation of this symmetric configuration holds:

$$\sigma_x = \frac{1}{2}L \tag{3.50}$$

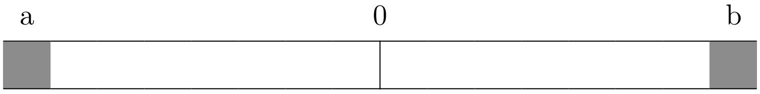


Table 3.3: Symmetrical periphery-heavy distribution

Intuitively, one would like to have the value L as the result of a measure of magnitude for this distribution especially. The factor 1/2 for the standard deviation or 1/4 for the variance has its reason firstly in the already above discussed factor 1/2 in Eqn. 3.46 and secondly in a second factor 1/2, which results from the fact that due to the accumulations at the outer boundaries also in the mean $M[(\Delta x)^2]$ as many differences $|x_i - x_j| = 0$ as $|x_i - x_j| = L$ are contained. $M[(\Delta x)^2]$ gives the expected result L in exactly one case, namely when $N = 2$, which excludes differences $|x_i - x_j| = 0$. Overall, the results reflect nothing more than the simple fact that the mean of the distance squares is smaller than their maximum. On the other hand, this makes it very clear that the definition of the system size in the notion of extensive spreading is not unique, for there is no reason why one definition should in principle be given precedence over the other. This becomes even clearer if, instead of the periphery-heavy distribution, one looks at the center-heavy distribution as in table 3.4 below, which in the extreme case has exactly one function value at each of the two outer boundaries, but $x_i = 0$ holds for all other i, with which the variance disappears for large N.

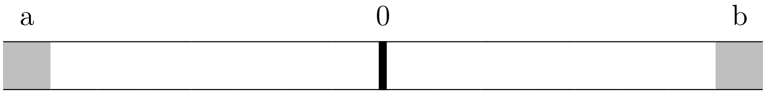


Table 3.4: Center-heavy distribution

Between these two extreme cases, another characteristic distribution is the uniform distribution, represented by the following table:

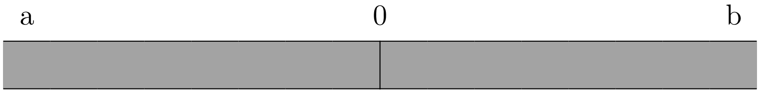


Table 3.5: Uniform distribution

To calculate its variance, we can first note that the distribution must be symmetric in relation to the reference point. Moreover, the sum $\sum_i x_i^2$ is independent of the order of x_i. Therefore, we can assume that $i < j \Leftrightarrow x_i < x_j$ holds. If we further replace N by $2N + 1$ for simplicity, then it follows according to Eqn. 3.48 together with $\Delta x := x_{i+1} - x_i$ which is independent of i – because of the uniform distribution:

$$\sigma_x^2 = \frac{1}{2N+1} 2 \sum_{i=1}^{N} (\Delta x \cdot i)^2 = \frac{2(\Delta x)^2}{2N+1} \frac{N(N+1)(2N+1)}{6} \approx \frac{1}{12}(2N\Delta x)^2 \approx \frac{1}{12}L^2 \tag{3.51}$$

This results in the standard deviation of the uniform distribution for large N:

$$\sigma_x \approx \frac{1}{\sqrt{3}} \frac{1}{2} L \tag{3.52}$$

The result places the standard deviation of the uniform distribution in a fixed relation to the maximum spread L of the system. Although this is also true for the periphery-heavy distribution given by Eqn. 3.50. However, in terms of extensivity, the uniform distribution is unique. This is because a uniformly distributed system S with sizes L, N can be arbitrarily decomposed into sufficiently large, again uniformly distributed subsystems S_1 and S_2 with sizes L_1, N_1 and L_2, N_2, respectively, for whose standard deviation then holds:

$$\sigma_x(S) \approx \frac{1}{\sqrt{3}} \frac{1}{2} L = \frac{1}{\sqrt{3}} \frac{1}{2} (L_1 + L_2) \approx \sigma_x(S_1) + \sigma_x(S_2) \tag{3.53}$$

The relation states that the standard deviation is more or less good within a uniformly distributed system depending on the size of N, but strictly extensive in the limit $N \to \infty$ for all subsystems.

Extensivity in continuous systems

The end of the last section with Eqn. 3.51 for the uniform distribution already marks the transition from finite to continuous systems, which is also equivalent to the transition from *Riemann sums*[14] of the square function to the integral with value $(1/12)L^2$, as is shown below in Eqn. 3.54, because for the limit of the left-hand side of Eqn. 3.51 we get because of $L = (2N+1)\Delta x$:

$$\lim_{N\to\infty} \left(\frac{1}{2N+1} 2 \sum_{i=1}^{N} (\Delta x \cdot i)^2 \right) = \lim_{N\to\infty} \left(\frac{2}{L} \sum_{i=1}^{N} i^2 \cdot \Delta x \right) = \frac{1}{L} \int_{-L/2}^{+L/2} x^2 \, dx = \frac{1}{12} L^2 \tag{3.54}$$

For general distributions, the change from finite discrete to continuous distributions takes place in three steps. The first step consists in the combination of function values x_i with the same value. I.e. the function

$$x : i \mapsto x_i$$

is given a second function

$$p : i \mapsto p_i = \frac{n_i}{N} \tag{3.55}$$

with the frequencies n_i of the function value x_i or its relative frequencies p_i. This step is simply an equivalent representation of the same facts. The mapping x thus becomes an injective function. The variance according to Eqn. 3.48 becomes:

$$\sigma_x^2 = \frac{1}{N} \sum_i n_i x_i^2 = \sum_i p_i x_i^2 \tag{3.56}$$

---

[14] https://en.wikipedia.org/wiki/Riemann_sum

The second step is based on the fact that the function x is now, because injective, also dispensable. The original function $x : \mathcal{M} \to \mathbb{R}$ is a namespace in the sense of section 3.3.1 for distinguishing the elements of a set which do not appear in the function itself. In statistics, for example, the elements of this set may be numbered individual experiments with the same result, so that the experiments cannot be distinguished by the results, but can be distinguished by the index i. The injectivity of x therefore means that the elements of the set $x(\mathcal{M})$ are distinguishable – now self-referentially – even without reference to the namespace. The index set $\mathcal{M}$ is thus represented by its image set $\mathcal{M}' := x(\mathcal{M}) \subseteq \mathbb{R}$, and we are left with the function:

$$p : \mathcal{M}' \to \mathbb{R}, x \to p(x)$$

The third step is to turn the finite set $\mathcal{M}'$ into an uncountable set $\mathcal{C} \subseteq \mathbb{R}$ of the continuum and the probability distribution p into the probability density P:

$$P : \mathcal{C} \to \mathbb{R}, x \to P(x)$$

Thus, the variance accourding to Eqn. 3.56 becomes:

$$\sigma_x^2 = \int_{\mathcal{C}} P(x) x^2 \, dx \tag{3.57}$$

From this, the variance of the uniform distribution with the constant function $P(x) \equiv c$ on the interval $\mathcal{C} = [-L/2, L/2]$, for which follows from the normalization condition

$$1 = \int_{\mathcal{C}} P(x) \, dx = \int_{-L/2}^{L/2} c \, dx = cL$$

$c = 1/L$, yields the same integral as in Eq. 3.54 and therefore the same result.

A periphery-heavy continuous distribution on the interval $[-L/2, L/2]$ can be represented by the uniform distribution on the set $\mathcal{C} = [-L/2, -L/2 + l] \cup [L/2 - l, L/2]$ with $P(x) = 1/2l$ and $l \ll L$. For their variance we get

$$\sigma_x^2 = \frac{1}{2l} \cdot 2 \int_{(L/2)-l}^{L/2} x^2 \, dx \approx \frac{1}{l} \left(\frac{1}{3} 3 \left(\frac{L}{2} \right)^2 l \right) = \frac{1}{4} L^2$$

thus also the same result as in the previous section for finite discrete systems, for which, however, the results are exactly valid only if the limit $N \to \infty$ is taken, which also applies to Eqn. 3.53, i.e., for the extensivity of the discrete uniform distribution. Therefore, we can assume that the continuity of the distribution is a prerequisite for the standard deviation to be guaranteed extensive without error. And more precisely, as already indicated by the discrete uniform distribution, it is the continuous uniform distributions that satisfy extensivity with precision, which becomes clear in the following way:

Let

$$P : [a, b] \to P(x)$$

be any distribution on the interval $[a, b]$ with $a < 0 < b$, for which 0 is not necessarily the distribution mean, but any point in the interval. Then P on each subinterval $[0, L] \subseteq [a, b]$ defines a distribution $Q_L : x \mapsto Q_L(x)$ which emerges from P by re-normalization. The distribution P is now extensive if and only if the standard deviations $\sigma(Q_L)$ of the distributions Q_L grow in the same way as the lengths L of the intervals $[0, L]$, i.e., if $\sigma(Q_L)$ depends linearly on L:

$$\sigma(Q_L) = \lambda L \tag{3.58}$$

With the definition equation 3.44 or its continuous version this means:

$$\int_0^L Q_L(x)x^2\,dx - \left(\int_0^L Q_L(x)x\,dx\right)^2 = \lambda^2 L^2 \tag{3.59}$$

This integral equation for $Q_L(x)$ in the variables x and L is solved by the uniform distributions $Q_L \equiv \frac{1}{L}$, which is shown by substituting the function into the equation, which yields $\lambda^2 = 1/12$, from which we finally obtain as in Eqn. 3.54 for all distributions Q_L:

$$\sigma(Q_L) = \frac{1}{\sqrt{12}}L \tag{3.60}$$

Intensivity

Here we want to investigate the sensitivity with which the standard deviation is able to map contractions or expansions of the distribution inside the system. To do so, we consider local but – to register anything at all – system-wide expansions of a discrete system of length L with high resolution N and their variance according to Eqn. 3.56. Let $p_i x_i^2$ be one of the N summands of the variance and $\Delta x_i = x_{i+1} - x_i$. Let Δp_i be a part of p_i shifted to $x_i + \Delta x_i/2$. Then, if Δx_i is very small, the proportion of this pointwise shift to the total change in variance is equal to:

$$\left((p_i - \Delta p_i)x_i^2 + \Delta p_i\left(x_i + \frac{\Delta x_i}{2}\right)^2\right) - p_i x_i^2 \approx \Delta p_i \cdot \Delta x_i \cdot x_i$$

When the limit $N \to \infty$ is taken this change turns into a differential $x \cdot dx \cdot dp$ of second order whose – also system-wide – integral always vanishes.

It follows that the variance practically does not register contractions or expansions on small scales, so it does not reflect the intensive size of a system either.

Counting variance

If we have established now that the standard deviation perceives internal contractions only insufficiently, then there is nevertheless a possibility to make it also receptive to it. And this can be done by its adaptation to the hierarchical, i.e. the intensive structure of space, which is already illustrated by table 2.6 with recursive bisection of the system cells. One can then introduce a local scale on each bisected cell and assign to its two halves a -1 and $+1$, respectively, as we are used to doing with a scale, which can be seen in the following figure:

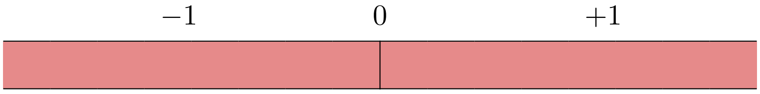


Table 3.6: Scale to distinguish the two halves of a cell

We then proceed, as at the beginning of section 3.5.1, first from a distribution:

$$x : \mathcal{M} \to \mathbb{R}, i \mapsto x_i$$

To make it concrete, let us assume that the mapping x describes the states of classical particles in one-dimensional position space, from which we now pick out those residing in the bisected cell. Let the set $\mathcal{M}$ thus be a finite set of indices for distinguishing the N particles in the bisected cell. The mapping x then becomes the mapping

$$x : \mathcal{M} \to \{-1, +1\}, i \mapsto x_i$$

We could then calculate the variance of the distribution according to the mean expression of Eq. 3.44:

$$\sigma_x^2 = \frac{1}{N}\sum_i x_i^2 - \left(\frac{1}{N}\sum_i x_i\right)^2$$

While the first sum $\sum_i x_i^2 = N$ always counts all N particles, for the second sum $\sum_i x_i$ holds the following:

$$\sum_i x_i = \sum_{x_i=-1} x_i + \sum_{x_i=+1} x_i = -\sum_{x_i=-1} (-x_i) + \sum_{x_i=+1} x_i =: -N_- + N_+$$

N_- counts the particles of the left half of the cell and N_+ correspondingly those of the right half. Thus, the sum $\sum_i x_i = N_+ - N_-$ establishes the difference of the particle numbers of both halves. With the relative frequencies $p_- := N_-/N$ and $p_+ := N_+/N$ follows for the variance:

$$\sigma_x^2 = 1 - \left(\frac{N_+ - N_-}{N}\right)^2 = 1 - (p_+ - p_-)^2 = (1 - p_+ + p_-)(1 + p_+ - p_-) = 4p_-p_+$$

The standard deviation is thus determined by the geometric mean of the proportions of the two halves:

$$\sigma_x = 2\sqrt{p_-p_+} \tag{3.61}$$

Their behavior can be represented with $x := p_-$ by the function $x \mapsto 2\sqrt{x(1-x)}$. The following figure shows its graph in blue along with the Shannon entropy $x \mapsto -x\log(x) - (1-x)\log(1-x)$ in yellow:

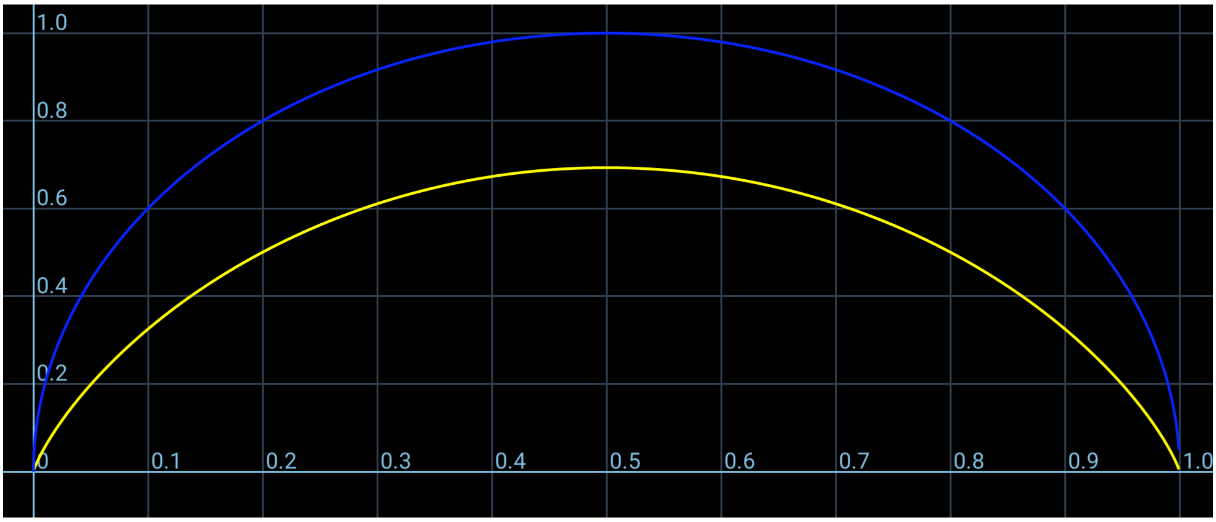


Figure 3.6: Standard deviation according to Eqn. 3.61 together with the Shannon entropy

This standard deviation thus favors uniform distribution in the same way as Shannon entropy, and also like Shannon entropy can grow at any scale if, for example – as with Shannon entropy – the respective increase by halving a cell is weighted by its own relative size.

3.5.2 Shannon entropy and its self-reference

Problem of extensivity and the counting principle

In the introduction we presented the model of a system for which the counting principle, or more precisely the product rule of the principle, also defined in the introduction, has no validity. Now the extensivity of the Gibbs entropy and that of the formally equal Shannon entropy stands or falls with the product rule. For the Boltzmann entropies or the Shannon entropies of uniformly distributed systems this follows directly from the additivity of the logarithm. For the Shannon entropies of two systems with arbitrary distributions $p_{11}, p_{12}, p_{13} \cdots p_{1m}$ and $p_{21}, p_{22}, p_{23} \cdots p_{2n}$ holds:

$$-\sum_i p_{1i} \log(p_{1i}) - \sum_j p_{2j} \log(p_{2j}) = -\sum_{ij} p_{1i} p_{2j} \log(p_{1i} p_{2j}) \tag{3.62}$$

Thus, Shannon entropy is additive exactly when the probability products $p_{1i} \cdot p_{2j}$ are the probabilities of the composite system. With reference to this, in the following I will give – in addition to the model in the introduction – further reasons why the Shannon entropy cannot be a physical entropy:

Independence: The systems under consideration are not systems with separate subsystems. Rather, we are concerned with spatially adjacent subsystems whose independence is difficult to establish. In any case, the independence is of unknown magnitude. Above all, it can be assumed that with continued partitioning of the system into increasingly small subsystems, the dependence increases, so that extensivity is valid at best in the limiting case of large subsystems.

External reference: If we consider Shannon entropy as a model for density entropy, then it has no reference to the size of space nor to the size of the absolute number of particles. The absence of this reference makes its definition self-referential, because the probability distribution $p_1, p_2, p_3 \cdots$ fundamentally lacks reference to an environment. This is because the probabilities, which we also interpret as *relative quantities*, map exclusively internal relations. The extensive quantity of a system, however, is an external relation with respect to the rest of the universe, as can be seen following Eqn. 3.38 for the density-entropy, which – in the mentioned equation without absolute number of particles – has also only logarithmic extensivity, but which thus at least stands in a relation to the extensive size of space at all. In contrast to this, Shannon entropy is the exclusive result of a self-description of the system.

Extensive connection: The addition of the entropies in Eqn. 3.62 is not the only way of combining two distributions $p_{11}, p_{12}, p_{13} \cdots p_{1m}$ and $p_{21}, p_{22}, p_{23} \cdots p_{2n}$ in the systematics of Shannon entropy. On the contrary, since the context of extensivity is about linking distributions of different systems, Eqn. 3.13 gives the following impetus to unify the two distributions: to this end, let p_1 and p_2 be the relative sizes of the two systems, then according to Eqn. 3.13, the following sum is the Shannon entropy of the composite system:

$$p_1 \left(-\log(p_1) - \sum_i p_{1i} \log(p_{1i}) \right) + p_2 \left(-\log(p_2) - \sum_j p_{2j} \log(p_{2j}) \right) \tag{3.63}$$

Unlike in Eqn. 3.62, however, no products $p_{1i} \cdot p_{2j}$ arise here, so the question of the validity of the product rule of the counting principle does not even arise. Rather, the Shannon entropy is then certainly not additive, because the sum in 3.63 does not agree with the left side of Eqn. 3.62.

Complementarily, one can check, as it is done in the following section on density entropy, whether the Shannon entropy complemented with the particle number N, i.e. the entropy $-N\sum p_i\log(p_i)$ from the introduction or according to Eqn. 3.7 is additive. For the sum of two such entropies $S_1 = -N_1\sum p_{1i}\log(p_{1i})$ and $S_2 = -N_2\sum p_{2j}\log(p_{1j})$ is obtained with probabilities $p_1 := N_1/(N_1+N_2)$ and $p_2 := N_2/(N_1+N_2)$:

$$
\begin{aligned}
S_1+S_2 &= N_1\left(-\sum_i p_{1i}\log(p_{1i})\right)+N_2\left(-\sum_j p_{2j}\log(p_{2j})\right)\\
&=(N_1+N_2)\left(\frac{N_1}{N_1+N_2}\left(-\sum_i p_{1i}\log(p_{1i})\right)+\frac{N_2}{N_1+N_2}\left(-\sum_j p_{2j}\log(p_{2j})\right)\right)\\
&=(N_1+N_2)\left(p_1\left(-\sum_i p_{1i}\log(p_{1i})\right)+p_2\left(-\sum_j p_{2j}\log(p_{2j})\right)\right)
\end{aligned}
$$

It can be seen from the last line that although the sum is correlated in magnitude with and only with the number of particles, because the content of the large parenthesis is a mean value. On the other hand, from the comparison with Eqn. 3.63 it is clear that the content of this parenthesis does not have the form of Shannon entropy, i.e., it lacks, compared to this one, the entropy

$$-p_1\log(p_1)-p_2\log(p_2)$$

resulting from the bipartition, so that the defined entropy $-N\sum p_i\log(p_i)$ grows more by system expansion than the entropy sum S_1+S_2.

Intensivity

In the case of the interpretation of Shannon entropy as an intensive quantity, we combine the two probability distributions and their product on the right-hand side of Eqn. 3.62, in which one distribution, e.g., $p_{21}, p_{22}, p_{23}\cdots p_{2n}$, the same decomposition $p_{1i}\cdot p_{21}, p_{1i}\cdot p_{22}, p_{1i}\cdot p_{23}\cdots p_{1i}\cdot p_{2n}$ of each probability p_{1i} of the other distribution describes, with a hierarchically deepened and system-wide uniform spatial resolution of a *single* physical system. (In the extensive interpretation of Eqn. 3.62, on the other hand, each of the two distributions describes the resolution of the state space of a separate physical system each.)

To assess whether Shannon entropy is an intensive measure, it should be possible to distinguish extensive from intensive change[15]. However, insofar as one gives preference to the resolution, i.e., understands changes as intensive and not as extensive change, the Shannon entropy according to section 3.3.3 even increases infinitely and insofar maps the resolution after all, but at the same time is indefinite precisely because of its infinity. This problem in turn is obviously met by the creation of a concept, the *coarse graining*, as the following quotation shows:

> *In classical statistical mechanics, the number of microstates is actually uncountably infinite, since the properties of classical systems are continuous. For example, a microstate of a classical ideal gas is specified by the positions and momenta of all the atoms, which range continuously over the real numbers. If we want to define Ω, we have to come up with a method of grouping the microstates together to obtain a countable set. This procedure is known as coarse graining. In the case of the ideal gas, we count two states of an atom as the "same" state if their positions and momenta*

---

[15] Cf. the end of section 3.3.2

> *are within δx and δp of each other. Since the values of δx and δp can be chosen arbitrarily, the entropy is not uniquely defined. It is defined only up to an additive constant.*[4]

Coarse graining says nothing else than that the resolution is limited because the entropy definition opposes infinitely fine resolution. Thus, its value depends on the arbitrarily chosen limit. Finally, the statement, also made in the quotation, that regardless of this dependence, entropy is well defined except for a constant, is certainly wrong, because it assumes that below a certain scale the distinction between uniform and non-uniform distribution is meaningless, but for the scales above it is not. I.e., actually it is assumed that below this scale there is only uniform distribution, in which case the extension of the resolution and the entropy increase connected with it are only caused by the continued division and therefore appear as meaningless. As will be shown to some extent already in the following section, however, it is just the other way round, namely that taking the limit to infinitesimal units is exactly one of the essences of physical entropy.

3.5.3 Density entropy and its double reference

Extensivity

The density entropy according to Eqn. 3.38 reflects the dualism of space and matter. In this respect, the density entropy is a double-referential quantity measure in contrast to the Shannon entropy. Despite this connection, however, density entropy also requires a completion to be extensive. In equation 3.35 it is defined as the sum of Shannon entropy and an averaged complexity of the continuum, which, as for Shannon entropy alone, we cannot expect to be extensive. Therefore, as was done for the Shannon entropy in section 3.5.2 in a supplementary experiment, we also consider the particle number N. From the defining equation 3.35 we then get

$$S_D(N) := -N\sum_i p_i \log(p_i) + N\sum_i p_i \log\left(\frac{\Delta x_i}{u}\right) \tag{3.64}$$

and from Eqn. 3.38:

$$S_D\left(N, \frac{L}{u}\right) = N S_D\left(\frac{L}{u}\right) = N\log\left(\frac{L}{u}\right) - N\sum_i p_i \log\left(\frac{p_i}{q_i}\right) \tag{3.65}$$

By considering the number of particles now the density entropy becomes an additive and therefore also extensive quantity. To show this, with definition of the relative lengths $l_1 := L_1/u_1$ and $l_2 := L_2/u_2$, let

$$S_1 = N_1 \log(l_1) - N_1 \sum_i p_{1i} \log\left(\frac{p_{1i}}{q_{1i}}\right)$$

$$S_2 = N_2 \log(l_2) - N_2 \sum_j p_{2j} \log\left(\frac{p_{2j}}{q_{2j}}\right)$$

be the density entropies of two neighboring systems. With $N := N_1 + N_2$, $p_1 := N_1/N$, $p_2 := N_2/N$ and two initially arbitrary probabilities q_1, q_2 with $q_1 + q_2 = 1$ we obtain:

$$S_1+S_2 = N\left(p_1\log(l_1) - p_1\sum_i p_{1i}\log\left(\frac{p_{1i}}{q_{1i}}\right) + p_2\log(l_2) - p_2\sum_j p_{2j}\log\left(\frac{p_{2j}}{q_{2j}}\right)\right) \tag{3.66}$$

$$= N\left(p_1\log\left(\frac{l_1p_1}{q_1}\right) + p_2\log\left(\frac{l_2p_2}{q_2}\right) - \sum_i p_1p_{1i}\log\left(\frac{p_1p_{1i}}{q_1q_{1i}}\right) - \sum_j p_2p_{2j}\log\left(\frac{p_2p_{2j}}{q_2q_{2j}}\right)\right) \tag{3.67}$$

Now the two product sequences $p_1p_{11}, p_1p_{12}, \cdots$ and $p_2p_{21}, p_2p_{22}, \cdots$ together form again a probability distribution and so do the sequences $q_1q_{11}, q_1q_{12}, \cdots$ and $q_2q_{21}, q_2q_{22}, \cdots$ so that the last expression above has the form of a density entropy according to Eqn. 3.65 if there is a relative length l for the composite system and values for the unknowns q_1 and q_2 such that

$$p_1\log\left(\frac{l_1p_1}{q_1}\right) + p_2\log\left(\frac{l_2p_2}{q_2}\right) = \log(l) = p_1\log(l) + p_2\log(l)$$

applies. This is certainly the case when

$$\frac{l_1p_1}{q_1} = \frac{l_2p_2}{q_2}$$

is fulfilled, from which together with $q_1 + q_2 = 1$ at first

$$q_1 = \frac{p_1l_1}{p_1l_1 + p_2l_2}, q_2 = \frac{p_2l_2}{p_1l_1 + p_2l_2} \tag{3.68}$$

follows and with $L/u := l$ finally:

$$\frac{L}{u} = p_1\frac{L_1}{u_1} + p_2\frac{L_2}{u_2} \tag{3.69}$$

For the entropy sum we get

$$S_1+S_2 = N\left(\log\left(\frac{L}{u}\right) - \sum_i p_1p_{1i}\log\left(\frac{p_1p_{1i}}{q_1q_{1i}}\right) - \sum_j p_2p_{2j}\log\left(\frac{p_2p_{2j}}{q_2q_{2j}}\right)\right) \tag{3.70}$$

which therefore again, derived from the information of the subsystems, has the form of the density entropy 3.65, and which at the same time grows with the extensive particle number N and according to Eqn. 3.66 insofar with N *only* in that the content of the large parenthesis in this equation is the sum of two mean values, namely:

$$p_1\log(l_1) + p_2\log(l_2)$$

$$p_1\left(-\sum_i p_{1i}\log\left(\frac{p_{1i}}{q_{1i}}\right)\right) + p_2\left(-\sum_j p_{2j}\log\left(\frac{p_{2j}}{q_{2j}}\right)\right)$$

In the thermodynamic equilibrium of the composite system becomes from Eqn. 3.70, because then the intensive complexity $-\sum_i p_1 p_{1i} \log(\cdots) - \sum_j p_2 p_{2j} \log(\cdots)$ vanishes, together with Eqn. 3.69:

$$S_1 + S_2 = N \log \left(p_1 \frac{L_1}{u_1} + p_2 \frac{L_2}{u_2} \right) \tag{3.71}$$

As an example of the equation, we can consider the composition of two equal systems in thermodynamic equilibrium, both with particle number N, length $L^{'}$ and unit length $u^{'}$. For the extensive length L of the total system, $L = 2L^{'}$ holds. From Eqn. 3.71 it then follows because of $p_1 = p_2 = 1/2$

$$S_1 + S_2 = 2N \log \left(\frac{L^{'}}{u^{'}} \right) = 2N \log \left(\frac{2L^{'}}{2u^{'}} \right) = 2N \log \left(\frac{L}{2u^{'}} \right) \tag{3.72}$$

so that the system extension with $N \to 2N$ and $L^{'} \to L = 2L^{'}$ is associated with a scale transformation and transformation of the unit length

$$u^{'} \to u = 2u^{'}$$

One could conclude that the density entropy is also extensive only when the scale transformation is taken into account. However, the following explanation seems more reasonable, according to which the relative quantity L/u in Eqn. 3.65 has a hybrid meaning:

System composition: When systems are composed into a new whole, L/u is not an extensive quantity, which is clearly shown in Eqn. 3.69, because according to it the value for L/u of a composite system is the average of the single values L_i/u_i and therefore cannot correlate with this number when an increasing number of systems are united and thus cannot be extensive. But the quantity L is extensive. And if one now wants to express the entropy by this extensive, but actually inappropriate quantity, then the simultaneous adjustment of the scale is necessary accordingly. From this no arbitrariness is derived for the value of the entropy of the composite system, because the value for L/u is, after all, determined uniquely by Eqn. 3.69. Arbitrary, on the other hand, are the original quantities L_1/u_1 and L_2/u_2 of the initial systems, which is in the nature of the continuum, where there are no absolute values, but only relative ones. In this respect, the equations 3.69 and 3.71 are comparable with a logical implication.

System comparison: When comparing the quantities of different systems on the basis of a common reference scale and unit of measurement u, L/u has the meaning of an extensive quantity.

Furthermore, it becomes apparent how space and matter dualistically refer to each other. The double reference arises from the fact that the description of a system by its parts follows a different logic than that of the unification of whole systems to a new whole:

Description of a system by its parts: The description by division is about the description of a *single* system on the basis of its parts. Considering such a system, its first – external – description is the specification of its number of particles and that of its extensive size L in space in relation to a unit of measure u. The length of this description is the extensive entropy $N \log(L/u)$ given by Eqn. 3.65. For the further – efficient – description of the interior of the system, neither the particle number N, nor the quantities L and u are required. Rather, the continuation of the description consists exclusively in the determination of inner relations, so that for this purpose the particle number and the size of the space are both universally set to 1 and the inner relations are finally determined by dividing the system and pair the relative quantities p_i and q_i or determine the relative densities $\varrho_i = p_i/q_i$. In this composition, the definition of q_i precedes that of p_i. Thus, the relative particle numbers p_i are defined only in their reference to the relative cell sizes q_i. They are functionally dependent on the definition of q_i.

Description of a union of whole systems: The description of a system united from a number of whole systems, for the sake of simplicity only the unification of two, on the other hand has other conditions. Here, because it concerns whole systems, the complete descriptions of both systems are present by definition, thus in particular the absolute particle numbers N_1 and N_2, the extensive expansions L_1 and L_2 as well as the units u_1 and u_2. Now, first the relative particle shares $p_1 = N_1/(N_1 + N_2)$ and $p_2 = N_2/(N_1 + N_2)$ can be derived, but this time without reference to the relative quantities q_1 and q_2. Instead, it is the other way around, because now these are derived according to Eqn. 3.68 by reference from p_1 and p_2. Thus, the q_i are functionally dependent on the magnitudes of the p_i.

Intensivity

To study the intensivity of the density entropy, we consider, as for the Shannon entropy in section 3.3.3, the simple bisection of a subsystem with the relative particle share p_i and the relative cell size q_i to get an idea of how the density entropy behaves at increasing resolution. However, we fix the cell division by halving the cell size q_i and distribute the particle share p_i arbitrarily between the halves. If p_{i1} and p_{i2} are these new relative particle shares of the two halves of the cell relative to the total system with $p_{i1} + p_{i2} = p_i$, then the halving results in the following change ΔS_i in the entropy of the i-th cell:

$$\begin{aligned}
\Delta S_i &= -p_{i1}\log\left(\frac{p_{i1}}{q_i/2}\right) - p_{i2}\log\left(\frac{p_{i2}}{q_i/2}\right) + (p_{i1}+p_{i2})\log\left(\frac{p_i}{q_i}\right)\\
&= -p_{i1}\log(p_{i1}) - p_{i2}\log(p_{i2}) + (p_{i1}+p_{i2})\log(p_i) - p_i\log(2)\\
&= p_i\left(-\frac{p_{i1}}{p_i}\log\frac{p_{i1}}{p_i} - \frac{p_{i2}}{p_i}\log\frac{p_{i2}}{p_i}\right) - p_i\log(2)
\end{aligned}$$

The second line shows that the two components of the density entropy, the Shannon entropy and the complexity of the continuum, are independently affected by cell division. The third line shows that, except for the factor p_i, the change in the share of Shannon entropy again takes the form of Shannon entropy[16] and thus like it is non-negative. Thus, this share can at most increase upon cell division[17]. In contrast, the already negative complexity of the continuum continues to decrease, here by the amount $p_i\log(2)$. For the sum of increase and decrease holds:

The Shannon entropy S_{sh} in the large parenthesis of the third line satisfies the inequality[18]

$$0 \le S_{sh} \le \log(2)$$

from which follows by multiplying by p_i and subtracting $p_i\log(2)$

$$-p_i\log(2) \le p_i S_{sh} - p_i\log(2) \le 0$$

On the other hand, according to the above equation

$$\Delta S_i = p_i S_{sh} - p_i\log(2)$$

---

[16] Cf. Eq. 3.13

[17] Cf. Eq. 3.16

[18] Cf. Eq. 3.10 and 3.11

applies and therefore also

$$-p_i \log(2) \leq \Delta S_i \leq 0$$

which results in

$$-\log(2) \leq \Delta S \leq 0 \tag{3.73}$$

for the total entropy decrease $\Delta S = \sum_i \Delta S_i$ in case of globally performed cell division after summation over all cells.

Thus, like Shannon entropy, density entropy, regardless of the magnitude of p_i – and thus at any scale – has in principle the potential to change in the order log(2), but with the following differences:

Monotonicity at cell division: The entropy falls monotonically at cell division, but not *strongly* monotonically, which distinguishes it from Shannon entropy. Thus, the decrease in density entropy is not obligatory and instead depends in each divided cell on the distribution of matter between the two halves of the cell.

Specific nature of uniform distribution: The decrease is zero exactly when the particles are equally distributed on both halves, while the increase of Shannon entropy has its maximum exactly in this case, because the increase itself has again the form of Shannon entropy maximized by the uniform distribution.

Convergence and divergence: While the Shannon entropy diverges like $\log(n)$ when divided depending on the scale n, the resolution, the density entropy has in principle the potential for both divergence and convergence.

Completeness of definition: In summary, one can clearly see here the incompleteness in the definition of Gibbs entropy, which is no different from Shannon entropy, namely with respect to its meaning in position space or momentum space. For the quantity p_i, with which the statement is associated that the relative particle share has just this value p_i, receives a meaning only by its relation to the assigned quantity q_i in the respective space in the form of the density p_i/q_i. Without this relation the definition of p_i is arbitrary and finally meaningless.

3.5.4 Relationship between extensive and intensive size

The aim of this section is to highlight the relationship between extensive and intensive size. For this purpose, we assume that the standard deviation or, if possible, the volume within fixed boundaries in space are suitable measures for the extensive quantity. The advantage of the standard deviation is its greater generality, since it allows a finite, i.e., definite, measure even for systems with infinite spread. Its disadvantage is that it is not itself strictly extensive, i.e. additive.

As the previous sections have shown, the standard deviation is not sensitive to the relationship between expansion and contraction inside the system. In fact, the volume excludes the perception of internal changes altogether. In contrast, density entropy, as per the previous section, is sensitive to both changes relative to the external environment and expansions or contractions occurring inside the system. The density entropy can therefore be understood as a measure of the extensive and at the same time of the intensive system size. The relation between the extensive quantity, represented e.g. by the standard deviation σ, and the entropy S is the following (see below):

$$S \leq \log\left(\frac{\sigma}{u}\right) \tag{3.74}$$

As before, u denotes a fixed unit of measurement, which in the above inequality can also absorb a possible additive constant – and just as well the generally, however, fuzzy relation between the volume and the standard deviation as, e.g., according to Eqn. 3.60.

In the relation above, the uniform distribution of extensively finite systems and the normal distribution of extensively infinite systems each have a special meaning among all other distributions. Namely, they are precisely the distributions that maximize entropy with respect to a fixed extensive quantity expressed by the standard deviation. And for the entropy S_{max} of these two distributions holds:

$$S_{max} = \log\left(\frac{\sigma}{u}\right) \tag{3.75}$$

The entropy is therefore limited by the extensive measure upward and by this exactly then unambiguously determined if the matter is completely expanded in the inside. Every inner contraction on the other hand, i.e. deviation from the uniform distribution, is acknowledged with a deduction. The simplest conception is that distribution gaps – connected with contractions at other locations – lead to the entropy decrease. Then it is clear that volume in particular – as a measure of size – does not register the gaps. Entropy, however, does perceive them, and at any scale, so it responds not only to discrete density fluctuations but also to continuous ones. When the gaps – even gradually – disappear completely, the complete system size coincides with the external size, the logarithm of the volume.

It must be pointed out, however, that while the above relations 3.74 and 3.75 describe the relation of entropy S or S_{max} to the extensive size σ of the system, these entropies themselves do not have the property of extensivity because they are averages and, relatedly, not linear but only logarithmically additive, as shown in the preceding sections.

In the following we want to consider the described connection by the example of different entropies:

Shannon entropy: The relation also holds for Shannon entropy if one associates the number n of probabilities exclusively with the extensive size of the system, thus avoiding the problematic mixing with internal division processes. If one sets $u = 1$, the two relations above are nothing but Eqn. 3.10

$$S = -\sum_{i=1}^{n} p_i \log(p_i) \leq \log(n)$$

with the addition that the equality holds exactly in the case of uniform distribution.

Density entropy at finite extension: For the density entropy and differential entropy of systems with finite extension, it is Eqn. 3.36 and 3.43:

$$S \leq \log\left(\frac{L}{u}\right)$$

And again it is the uniform distribution which is distinguished by being the only function which maximizes the entropy.

Density entropy at unrestricted extension: For the normal distribution

$$P(x) = 2\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\dfrac{x^2}{2\sigma^2}}$$

on the interval $[0,\infty)$ and the differential entropy according to Eqn. 3.40

$$S = -\log(u) - \int_0^\infty P(x)\log(P(x))\,dx$$

we obtain with

$$\sigma^2 = \int_0^\infty P(x)x^2\,dx$$

$$S = -\log(u) - \int_0^\infty P(x)\left(\frac{1}{2}\log(2) - \frac{1}{2}\log(\pi\sigma^2) - \frac{x^2}{2\sigma^2}\right) = \log\left(\frac{\sigma}{u}\right) + \log\left(\sqrt{\frac{e\pi}{2}}\right) \tag{3.76}$$

However, since the normal distribution not only exactly satisfies this equation, but – for a fixed standard deviation – also maximizes the *differential entropy*[19], the two relations 3.74 and 3.75 for the entropy above also apply here.

3.5.5 Relationship between uniform distribution and normal distribution

The last section suggests a close relationship between the uniform distribution and the normal distribution, which will be explored in the following sections.

Norm conserving expansion and contraction transformations

Let $\mathcal{F}_1(L)$ be the function space of functions bounded and normalized on the finite interval $[0, L]$. For a function

$$f : [0, L] \to \mathbb{R}_+$$

then $f \in \mathcal{F}_1(L)$ holds exactly if it is bounded and integrable and if at the same time

$$\int_0^L f(x)\,dx = 1$$

is true. Now let

$$\mathcal{F}_1 := \bigcup_{L\in(0,\infty)} \mathcal{F}_1(L)$$

be the set of all functions normalized respectively on a finite interval $[0, L]$.

In parallel, we define the function space $\mathcal{F}_2$ of the functions $f : [0,\infty) \to \mathbb{R}_+$ restricted to $[0,\infty)$ and normalized, for which thus also holds:

$$\int_0^\infty f(x)\,dx = 1$$

Finally, let

$$\mathcal{F} := \mathcal{F}_1 \cup \mathcal{F}_2$$

be the set of all these normalized functions with the bijective mappings

$$E_a : \mathcal{F} \to \mathcal{F}, f \mapsto f_a := E_a(f) \tag{3.77}$$

for $a \in (0,\infty)$, which is given by the equation:

$$f_a(x) = \frac{1}{a} f\left(\frac{x}{a}\right) \tag{3.78}$$

---

[19] https://en.wikipedia.org/wiki/Differential_entropy#Maximization_in_the_normal_distribution

For $f \in \mathcal{F}_1(L)$ f_a is defined on $[0, aL]$. Correspondingly, with $f \in \mathcal{F}_2$, f_a is also a function on $[0, \infty)$. Further, in both cases, i.e. if $f \in \mathcal{F}$, f_a is certainly bounded and integrable, but also normalized, because compared to f the domain of f_a, if e.g. $a > 1$, is stretched by a factor a and at the same time the image is compressed by $1/a$. More precisely, the normalization follows from *integration by substitution*, because then holds:

$$\int_0^{aL} f_a(x)\,dx = \int_0^{aL} \frac{1}{a} f\left(\frac{x}{a}\right)\,dx = \int_0^{L} f(x)\,dx = 1$$

Correspondingly, for $f \in \mathcal{F}_2$:[20]

$$\int_0^{\infty} f_a(x)\,dx = \int_0^{\infty} \frac{1}{a} f\left(\frac{x}{a}\right)\,dx = \int_0^{\infty} f(x)\,dx = 1$$

Finally, as claimed, E_a is also bijective, since $E_{1/a}$ is its inverse function.

Moreover, for all functions $f \in \mathcal{F}$ holds:

$$\lim_{a \to \infty} f_a(x) \equiv 0 \tag{3.79}$$

Namely, if max_f is the upper bound of the function $f \in \mathcal{F}$, then $f_a(x) \leq (1/a) max_f$.

Since all functions in $\mathcal{F}$ are normalized, they can also be interpreted as distributions. For $a > 0$ the mapping E_a is then an expansion transformation and for $a < 0$ a contraction. The following two figures show the relationship between contracted and expanded normal distribution and uniform distribution, respectively:

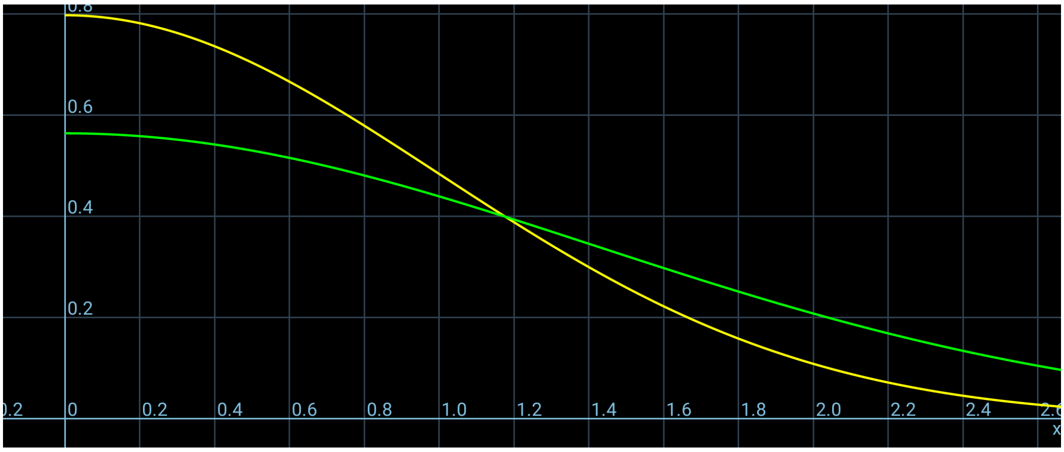


Figure 3.7: Normal distribution (yellow) and its expansion with factor $\sqrt{2}$ (green)

---

[20] Cf. section 2.3.1 on time reversal and reversibility of the heat equation

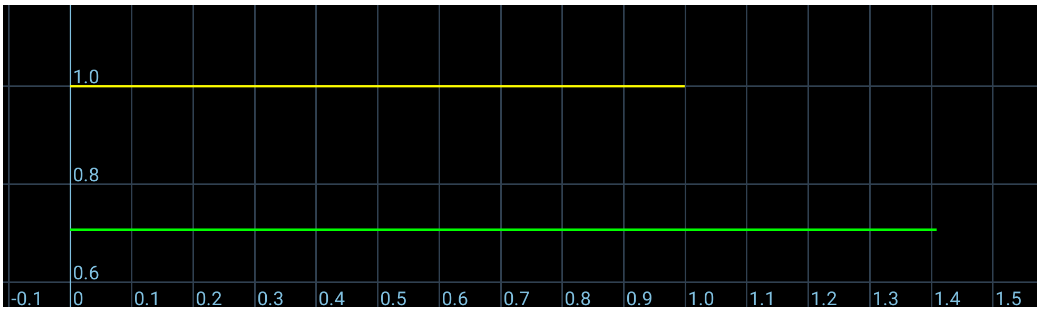


Figure 3.8: Uniform distribution (yellow) and its expansion with factor $\sqrt{2}$ (green)

Physically, conserving the norm corresponds to a conservation law, specifically that of the number of particles or energy.

For the differential entropy $S(f_a)$ of the transformed distribution f_a of a distribution $f \in \mathcal{F}_1(L)$ applies

$$\begin{aligned}
S(f_a) &= -\log(u) - \int_0^{aL} f_a(x)\log(f_a(x))\,dx \\
&= -\log(u) - \int_0^{aL} \frac{1}{a} f\left(\frac{x}{a}\right) \log\left(\frac{1}{a} f\left(\frac{x}{a}\right)\right)\,dx \\
&= -\log(u) - \int_0^{aL} \frac{1}{a} f\left(\frac{x}{a}\right) \log f\left(\frac{x}{a}\right)\,dx + \log(a)\int_0^{aL} \frac{1}{a} f\left(\frac{x}{a}\right)\,dx \\
&= -\log(u) - \int_0^{L} f(x)\log(f(x))\,dx + \log(a)\int_0^{L} f(x)\,dx
\end{aligned}$$

and so:

$$S(f_a) = S(f) + \log(a) \tag{3.80}$$

The same result is obtained for distributions $f \in \mathcal{F}_2$. Thus, the entropy increase $\log(a)$ associated with the expansion is exclusively a function of the extensive spreading and, in particular, independent of the function f.

Correspondingly, for the standard deviation σ of the distributions $f \in \mathcal{F}$ follows

$$\sigma(f_a) = a \cdot \sigma(f) \tag{3.81}$$

which in this respect therefore resembles the length L of the domains $[0, L]$ of the $\mathcal{F}_1$-functions. For it holds exemplarily for $f \in \mathcal{F}_2$:

$$\begin{aligned}
\sigma^2(f_a) &= \int_0^{\infty} f_a(x)x^2\,dx - \left(\int_0^{\infty} f_a(x)x\right)^2 dx \\
&= \int_0^{\infty} \frac{1}{a} f\left(\frac{x}{a}\right) a^2 \left(\frac{x}{a}\right)^2 dx - \left(\int_0^{\infty} \frac{1}{a} f\left(\frac{x}{a}\right) a\frac{x}{a}\right)^2 dx \\
&= a^2 \int_0^{\infty} f(x)x^2\,dx - a^2\left(\int_0^{\infty} f(x)x\right)^2 dx \\
&= a^2\sigma^2(f)
\end{aligned}$$

Scale transformations

A scale transformation is the substitution of the unit of measurement first appearing in Eqn. 3.29. Suppose that u replaces the unit of measure $u^{'}$. Let the ratio of the two be:

$$a = \frac{u}{u^{'}} \tag{3.82}$$

If $g : x \mapsto g(x)$ is a distribution $g \in \mathcal{F}$ as in the previous section, then $x = x(u^{'})$ is a quantity relative to a pair consisting of the unit of measure and the reference scale in scale space defined jointly with it. With $x(u^{'}) \cdot u^{'}$ a point in space is marked relative to the reference point 0. For the same point to be marked on the scales of both units $u^{'}$ and u,

$$x(u) \cdot u = x(u^{'}) \cdot u^{'}$$

must be, thus also:

$$x(u) = \frac{x(u^{'})}{\frac{u}{u^{'}}} = \frac{x(u^{'})}{a}$$

The scale transformation together with the adjustment $u^{'} \to u = au^{'}$ of the units of measure thus implies the substitution

$$x(u^{'}) \to x(u) = \frac{x(u^{'})}{a}$$

For a distribution function $g : x \mapsto g(x)$, the scale transformation therefore implies that the argument x must be substituted in the same way.

If one additionally demands for the scale transformations that with them also the normalization is preserved, namely on every subinterval of the definition range of a function g, then they must finally be identical with the expansion and contraction transformations. For if f is the distribution arising from g by scale transformation and is normalized on each subinterval in the same way as g, then holds:

$$\int_0^{x/a} f(x^{'})\, dx^{'} = \int_0^{x} g(x^{'})\, dx^{'}$$

From this it follows by derivation

$$\frac{1}{a} f\left(\frac{x}{a}\right) = g(x)$$

and it holds inversely with $a^{'} = 1/a$

$$\frac{1}{a^{'}} g\left(\frac{x}{a^{'}}\right) = f(x)$$

so that the scale transformation with $a = u/u^{'} > 1$ corresponds to a contraction of the function g.

The limit connection

The sets $\mathcal{F}_1(L)$ form equivalence classes in $\mathcal{F}_1$. In each of these classes, by section 3.5.4, the uniform distributions with

$$u_L(x) \equiv \frac{1}{L}$$

are each distinguished by being the only ones of the functions normalized on $[0, L]$ that maximize differential entropy.

The set $\mathcal{F}_2$ can also be decomposed into equivalence classes by grouping in it functions each having the same standard deviation. Let $\mathcal{F}_2(\sigma) \subset \mathcal{F}_2$ be defined by:

$$f \in \mathcal{F}_2(\sigma) \Leftrightarrow \sigma(f) = \sigma \tag{3.83}$$

For each equivalence class $\mathcal{F}_2(\sigma)$, according to section 3.5.4, the normal distribution is

$$n_\sigma(x) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

is the only function that maximizes the differential entropy.

For $a \in (0, \infty)$ then holds:

$$E_a(\mathcal{F}_1(L)) = \mathcal{F}_1(aL) \tag{3.84}$$

$$E_a(\mathcal{F}_2(\sigma)) = \mathcal{F}_2(a\sigma) \tag{3.85}$$

Both equations follow from the defining equation 3.78 for the transformation E_a:

- If $f \in \mathcal{F}_1(L)$, then f is defined on $[0, L]$ and hence $E_a(f)$ is defined on $[0, aL]$, so $E_a(f) \in \mathcal{F}_1(aL)$. The equality in Eqn. 3.84 follows from the bijectivity of E_a.
- If $f \in \mathcal{F}_2(\sigma)$, then by definition f has standard deviation σ. According to Eq. 3.81, $E_a(f)$ thus has standard deviation $a\sigma$ and is therefore included in $\mathcal{F}_2(a\sigma)$. The equality in Eqn. 3.85 follows again from the bijectivity of E_a.

Equations 3.84 and 3.85 state that under the mapping E_a the image of each equivalence class of the sets $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, is in turn an equivalence class. In this context, the obvious interpretation of Eqn. 3.80 is the notion that E_a does not change the magnitude relations among the functions of an equivalence class if one considers the entropy difference of each two functions of a class as an expression of these relations. In other words, the entropy growth around the value $\log(a)$, which is the same for all functions, reflects exclusively the extensive expansion or contraction $L \to aL$ or $\sigma \to a\sigma$. Relations inside the function spaces, on the other hand, are conserved. The reason for this, however, is that also the relations between the function values of a function f are conserved, which already follows from the defining equation 3.78, because according to it:

$$\frac{f_a(x_1)}{f_a(x_2)} = \frac{f\left(\frac{x_1}{a}\right)}{f\left(\frac{x_2}{a}\right)}$$

Finally, we can define another bijective mapping A which pairs the equivalence classes of both sets $\mathcal{F}_1$ and $\mathcal{F}_2$. This can be done with any linear function that assigns a standard deviation σ to the length L of an interval $[0, L]$:

$$\sigma(L) = \lambda L$$

.

The mapping A then depends on λ and is defined as follows:

$$A_\lambda : \mathcal{F}_1(L) \mapsto \mathcal{F}_2(\sigma(L)) = \mathcal{F}_2(\lambda L) \tag{3.86}$$

In the following I consider the transformation E_a also as a function respectively on the power sets $\mathcal{P}(\mathcal{F}_1)$ and $\mathcal{P}(\mathcal{F}_2)$. On $\mathcal{P}(\mathcal{F}_1)$, for example, let E_a be the function

$$E_a : \mathcal{P}(\mathcal{F}_1) \to \mathcal{P}(\mathcal{F}_1), \mathcal{S} \mapsto E_a(\mathcal{S})$$

where $E_a(\mathcal{S})$ is the image set of $\mathcal{S}$ under the expansion and contraction transformation E_a as defined in 3.77. Thus, it follows from Eqn. 3.86 together with Enq. 3.84 and 3.85 initially:

$$\begin{aligned} A_\lambda \circ E_a(\mathcal{F}_1(L)) &= A_\lambda(\mathcal{F}_1(aL)) = \mathcal{F}_2(\lambda a L) \\ E_a \circ A_\lambda(\mathcal{F}_1(L)) &= E_a(\mathcal{F}_2(\lambda L)) = \mathcal{F}_2(a\lambda L) \end{aligned}$$

So for all $a > 0$ the following is true

$$A_\lambda \circ E_a(\mathcal{F}_1(L)) = E_a \circ A_\lambda(\mathcal{F}_1(L)) \tag{3.87}$$

and, because $E_{1/a}$ is the inverse of E_a, we get

$$E_{1/a} \circ A_\lambda \circ E_a(\mathcal{F}_1(L)) = A_\lambda(\mathcal{F}_1(L)) = \mathcal{F}_2(\lambda L) \tag{3.88}$$

from which by multiplication from left by $E_{1/a} \circ A_\lambda^{-1} \circ E_a$ also

$$E_{1/a} \circ A_\lambda^{-1} \circ E_a(\mathcal{F}_2(\lambda L)) = \mathcal{F}_1(L) \tag{3.89}$$

follows. Because the respective right-hand sides of the last two equations are independent of a, the limits for $a \to \infty$ exist on the left-hand side, yielding the following relationships between the normalized functions on the finite interval $[0, L]$ and those with standard deviation λL on the infinite interval $(0, \infty)$:

$$\lim_{a\to\infty} (E_{1/a} \circ A_\lambda \circ E_a(\mathcal{F}_1(L))) = \mathcal{F}_2(\lambda L) \tag{3.90}$$

$$\lim_{a\to\infty} (E_{1/a} \circ A_\lambda^{-1} \circ E_a(\mathcal{F}_2(\lambda L))) = \mathcal{F}_1(L) \tag{3.91}$$

Now these equations do not yet make any statement about the relation between the uniform distributions and the normal distributions, but only about that between the sets $\mathcal{F}_1(L)$ and $\mathcal{F}_2(\lambda L)$, which are assigned to each other by A_λ according to 3.86 – independently of E_a. Conversely, however, it follows that the equations also hold for subsets $\mathcal{S}_1 \subseteq \mathcal{F}_1(L)$ and $\mathcal{S}_2 \subseteq \mathcal{F}_2(\lambda L)$ if they are like those given by a mapping

$$B_\lambda : \mathcal{S}_1 \mapsto \mathcal{S}_2$$

mapped to each other, which is defined independently of E_a. This is certainly possible if there is a selection procedure for the elements of these subsets from the sets $\mathcal{F}_1(L)$ and $\mathcal{F}_2(\lambda L)$. Now, in any case, the definition of a choice is possible for the uniform distribution and the normal distribution due to their distinction by entropy maximization within the sets $\mathcal{F}_1(L)$ and $\mathcal{F}_2(\lambda L)$, respectively. Thus, we define the one-element subsets $\mathcal{S}_1(L) \subset \mathcal{F}_1(L)$ and $\mathcal{S}_2(\sigma) \subset \mathcal{F}_2(\sigma)$ such, that the first subset contains the uniform distribution u_L on the interval $[0, L]$ and the second contains the normal distribution n_σ with standard deviation $\sigma = \lambda L$. However, we can go one step further for simplicity and focus on the elements themselves instead of the one-element subsets. With this difference, analogous to A_λ in 3.86, we define the mapping B_λ as follows:

$$B_\lambda : u_L \mapsto n_\sigma = n_{\lambda \cdot L} \tag{3.92}$$

The above derivations, from the definition 3.86 to the taken limits, can then be performed analogously, so that the following relations hold for the relation between equal and normal distributions:

$$\lim_{a \to \infty} E_{1/a} \circ B_\lambda \circ E_a(u_L) = n_{\lambda \cdot L} \tag{3.93}$$

$$\lim_{a \to \infty} E_{1/a} \circ B_\lambda^{-1} \circ E_a(n_{\lambda \cdot L}) = u_L \tag{3.94}$$

Interpretation

From the mere assignment B_λ in Eqn. 3.92 between uniform and normal distributions, the equality relations of Eqn. 3.93 and 3.94 result. Because further the transformations E_a can also be interpreted as mere scale transformations, the distinguishability of the uniform distribution from the normal distribution is not guaranteed.

The two distributions differ essentially in that the normal distribution reflects the asymmetry of infinite space, but the uniform distribution does not reflect that of finite space. For even in a closed interval $[0, L]$ the points have different relations to each other. The peripheral points differ from the points in the interior of the interval. This difference is not reflected in any way in the function values of the distribution, which are the same everywhere, so that it symmetrizes the in and of itself asymmetric space. The normal distribution, on the other hand, by its function values, exposes the broken symmetry of space, which is related to a distinguished point at which it has its maximum.

If a function is uniformly distributed on every finite part of space, and it has a distinguished point, then it is normally distributed.

Chapter 4

Mass points, individuals and collectives

4.1 Extensive complexity of mass points

The definition of complexity in the previous chapters makes it difficult to reconcile it with the notion of classical physics of mass points, according to which each mass point has an exact location and an exact velocity

- in relation to a reference point resp. a reference velocity
- in a respective continuous space

For then the size of the state space as well as the description complexity of the position and the velocity of only one mass point is infinite, because the sets of its possible positions and its possible velocities are uncountable, which is reflected in the description complexity in the irrationality of the particle coordinates with infinitely many decimal places. Thus, for the sake of the finiteness of the complexity, one is forced to a discretization, i.e. to a limitation of its intensive size, the definition of which, however, is undetermined, which is, finally, one of the reasons why the continuum – in comparison with the discrete space – is given priority at all.

Now this is certainly true if one equates the complexity of the mass point with the complexity of its state space. The discretization of space and also the limitation of its extensive size is compelling. But even for a classical system that is finite in all respects, defining the particle complexity over its state space is difficult because the definition of the governing state space is indeterminate. For instance, if an otherwise free particle is confined in a space with only a few possible local states, say n, we are quick to identify the size of the governing state space for complexity with this number n, with the consequence that particle complexity is time-independent. If, on the other hand, the system size spans, say, several light-years and n is therefore correspondingly large, we would certainly not associate the complexity of a particle whose initial state we know with the number n at the very beginning of the observation. Rather, in this case we cannot avoid tying the particle complexity to the description of the actual particle state, which in turn means, conversely, especially for small systems, that the complexity is cyclic.

Accepting these problems, for one-dimensional systems one can define the complexity of a particle as an extensive description complexity of the actual state as follows based on Eqn. 3.29, e.g., for the state x of the particle in relation to a reference point in position space:

$$C := \log\left(\frac{|x|}{u}\right)$$

The attribute *extensive* is necessary because the definition excludes and must exclude the question of the precision of the distance $|x|$, i.e. that of its infinite intensive complexity.

Another problem of the definition is its dependence on the reference point, which does not even disappear if one is interested only in its changes instead of the complexity itself, which at least eliminates the dependence on the reference measure u. Namely, if x and $x+\Delta x$ are two different states of the particle, then when x changes to $x+\Delta x$:

$$\Delta C = \log\left(\left|\frac{x+\Delta x}{\Delta x}\right|\right)$$

The importance of the reference point is that by the common reference to it the description of a many-particle state is redundancy-free and thus in particular simple, but also contradiction-free. The complexity of the description of a many-particle state is in this way proportional only to the number of particles, which implies that the relative information between two particles, which is so important for the dynamics of the system, their relative positions in space and their relative velocities, is implicitly derived from the relations of both particles to the reference point. In contrast, the description of the many-particle state bypassing a reference point and instead based on all relative information is complex, the description effort proportional to the number of particle pairs and thus proportional to the square of the number of particles. Most importantly, the information is redundant with the risk of not matching and thus not uniquely defining a system. If one considers for example only three points in the plane, thus a triangle, then its definition is simple by indication of each point's position in relation to a common reference point, and the three positions will always unambiguously define a triangle. Relative information, on the other hand, such as only the distances between each two of the three points, must be available with error-free precision in order to uniquely define a triangle at all. On the other hand, the redundancy-free description is only seemingly unique if it does not include the information about actually existing lack of precision.

Now it is interesting that in his analysis of the many-particle system probably the most important decision of *Newton* is the decomposition of the system into the pairwise relations of two particles each, by which he made relative information the basis of his axiomatics and thus the basis of the dynamical evolution of a system, although he was convinced of the *Absolute Space*[1] in contrast to, for example, *Leibniz*, for whom only relative motion made sense, as is also evident from the article linked above. In the said Newton's decomposition a fundamental independence of the relation between each two particles from all other particles is expressed.

In a further attempt to define the complexity of a system of mass points, in order to avoid the problem of the reference point described above, we can, on the basis of the relative state of two particles with the relative distances $\Delta x, \Delta y$ and Δz, define their relative complexity in the continuous space of positions on the basis of Eqn. 3.29 and a reference length u as follows:

$$C := \log\left(\left|\frac{\Delta x \cdot \Delta y \cdot \Delta z}{u^3}\right|\right)$$

Here, however, an even more serious problem arises because, although the complexity does not depend on the choice of reference point, it does depend on the orientation of the reference axes in space and even diverges if at least one of the three components $\Delta x, \Delta y$ or Δz approaches zero.

Finally, the attempt to define the complexity as a pure distance complexity

$$C := \log\left(\frac{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}}{u}\right)$$

---

[1] https://en.wikipedia.org/wiki/Absolute_space_and_time

is insufficient because it does not take into account the complexity of the relative twists of the two-particle axes for a many-particle system.

Thus, one is again thrown back to a single space dimension for which, however, the two-particle complexity of the relative state with distance Δx can finally be defined:

$$C := \log\left(\frac{\Delta x}{u}\right) \tag{4.1}$$

This results in the change ΔC of complexity when changing the distance from Δx_1 to Δx_2:

$$C := \log\left(\frac{\Delta x_2}{\Delta x_1}\right) \tag{4.2}$$

4.2 From distinguishable individuals to collectives

The last section shows clearly that the definition of the complexity of a classical mass point or of the relation of two mass points is not possible if it considers the intensive description effort in addition to the extensive one. In addition, it is unsuitable for three-dimensional space. Now the pointed out difficulties are not actually limited to mass points. Rather, they arise from the assumption that particles are distinguishable individuals, which is why they are distinguished by an index in classical physics. So let us consider an individual particle with – in contrast to a mass point – non-vanishing size, i.e. a distribution in space which is possibly even unbounded, for instance in the sense of quantum mechanics. If this is not exactly the uniform distribution, then with regard to the size of its exact description in relation to a reference point and therefore for its complexity also exactly the same problems arise as in the case of the individual mass point. For it seems clear in any case that the description presupposes the selection – of a preferably distinguished – point of the distribution, e.g. the distribution mean, whose description in relation to a reference point, however, does not differ from that of a mass point. The same is true for the description of the relative state of two such distributions.

In summary: Individuals can be described in relation to a reference point of the continuous space and also in relation to each other only fuzzy. That is, either the description is indeterminate or the complexity is infinite. In particular, the definition is limited to one-dimensional space.

To continue, therefore, the restriction to mass points is again sufficient for the time being. The section 3.5.1 on the standard deviation gives a hint that the state description of the whole system is also possible in another way, as far as the distinguishability of the particles is renounced. Namely, this is done by introducing the relative frequency p_i in 3.55, with which particles are no longer distinguished, but only counted. Particles are then no longer distinguishable from others by virtue of themselves, but only by their state. Even more clearly the collectivization is shown in the counting variance in section 3.5.1, which again is related to entropy, if it is determined like the latter on all scales, in which case it uniquely determines the state of a system with many or even few particles – at the cost of particle individuality.

At this point it becomes clear that classical mechanics is a theory against the background of the extensive space, in which the infinity of the intensive space is blanked out and thereby the infinite effort of the state definition for an individual particle is just as little imposed, which however is necessary for a particle distinguishable from others, because this then also claims its individual description of its state.

4.3 From mass points to distributions

The surrender of the particle individuality mentioned in the last section is thus a necessary condition for the fact that the complexity of a physical system is finite. However, the collectivization of particles is in principle possible for mass points in the same way as for particles each with its own distribution. In this section, therefore, we will consider collectives of mass points in terms of both density entropy and Shannon entropy, regardless of the fact that Shannon entropy has already disqualified itself for a physical entropy due to its insufficient extensivity[2].

Namely, we start with the Shannon entropy, or the entropy

$$-N \sum_i p_i \log(p_i)$$

introduced in the introduction, and assume an infinite partitioning hierarchy to calculate the entropy. Because the number of particles N is finite, there exists a scale of the hierarchy such that for all cells C_i of the scale, either $p_i = 0$ or $p_i = 1/N$. The entropy at this scale is then equal to $N \log(N)$. The same is true for all other scales downscale, so the entropy limit exists at infinite resolution and has the value $N \log(N)$, which corresponds to the complexity defined in section 3.3.1 over finite ordered namespaces for distinguishing the N particles. Now this is a problem in that the entropy thus has a finite, but time independent value. Time dependence is only possible by limiting the resolution, which in turn is arbitrary and thus indeterminate.

The density entropy according to 3.64 is the sum of the entropy just considered and the complexity of the continuous space, which, except for the number of particles N, coincides with the mean cell complexity

$$\sum_i p_i \log\left(\frac{\Delta x_i}{u}\right)$$

but which diverges toward $-\infty$ in the limit of infinite resolution, because the cell sizes Δx_i converge to zero by definition. Since the first part of the density entropy has the finite value $N \log(N)$ at infinite resolution according to the above, the density entropy of a system of finitely many mass points is thus negatively infinite and thus indeterminate. It can have finite values only at finite resolution, in which case, however, like the Shannon entropy of a particle continuum, it is as indeterminate as the chosen resolution is.

In a summary to this point, with respect to the possibility of defining complexity for physical systems:

Particle individuals: Models based on particle individuals are ruled out because they contradict the intensive structure of space and therefore can at best represent extensive complexity, since each individual would have infinite complexity by itself.

Particle collectives: For models with collectives of mass points, the Shannon entropy is either constant in time, or it has a value that depends on the resolution and is thus indeterminate. The density entropy of such systems is either negatively infinite, or at – finite resolution – like the Shannon entropy, it has no definite value.

---

[2] cf. section 3.5.2 on the *extensivity* of Shannon entropy

Chapter 5

Monotonicity in classical physics

5.1 Correlation and monotonicity in position space

5.1.1 Free particles, Loschmidt's paradox and the arrow of Time

The approach to understanding entropy, interpreting it as a measure of a *molecular chaos*[1], and the approach of explaining the entropy increase with the *H-theorem*[2] by collisions between particles, met with criticism *Loschmidt's*, known as *Loschmidt's paradox*[3], according to which irreversibility contradicts the time reversibility (time symmetry) of the equations of motion, which means that the reverse evolution, in which a complex state evolves into a simple state, is also possible.

Now it follows from this argument only that among all the processes described by the equations of motion, for every process A there is a second process which – in the same time direction – reverses the evolution of A. It does not mean, on the other hand, that the process A reverses itself at a point in time. Rather, the reversal requires a preparation which mirrors all impulses of the system at a point in time, but which does not happen spontaneously and by itself. In this respect, the time symmetry of the equations of motion is not inconsistent with the monotonic change of a physical quantity in a process running by itself, which, as shown below, is certainly true for the simplest physical processes.

First of all, it is different with the *recurrence theorem*[4] of *Henri Poincaré*, whose simplified statement is that a physical system with restriction of the extensive many-particle phase space volume comes arbitrarily close to a state once taken arbitrarily often in the future. The condition of the restriction is:

> *A finite upper bound can be set on the total potentially accessible phase space volume. For a mechanical system, this bound can be provided by requiring that the system is contained in a bounded physical region of space (so that it cannot, for example, eject particles that never return) – combined with the conservation of energy, this locks the system into a finite region in phase space.*[10]

For an isolated system, the condition applies to momentum space with constant energy, but not to volume in position space. In this respect the recurrence theorem is related to Loschmidt's paradox. While the latter requires the one-time preparation of the system, the recurrence theorem

---

[1] https://en.wikipedia.org/wiki/Molecular_chaos

[2] https://en.wikipedia.org/wiki/H-theorem

[3] https://en.wikipedia.org/wiki/H-theorem#Loschmidt's_paradox

[4] https://en.wikipedia.org/wiki/Poincar%C3%A9_recurrence_theorem

presupposes the permanent external influence to prevent the expansion of the system in the position space.

In the following, we will show that the simplest physical systems – left to themselves – necessarily expand monotonically in position space, showing first that the condition of the restriction of the recurrence theorem is not physically guaranteed for isolated systems and second that these systems are characterized by at least one form of monotonicity, which is equivalent to the irreversibility of any state once assumed.

It is essential to note that the monotonicity is the compelling consequence of the deterministic equations of motion and does not contradict the untouched correlation of the particles. Thus, the equations of motion imply the monotonicity and at the same time conserve the correlation of the particles, which is ultimately Loschmidt's thesis, namely that the information of each state passed through is conserved in each future state and thus – in principle – reconstructible.

And namely, the monotonicity of state evolution is a consequence of the asymmetry of both extensive and intensive space. The asymmetry of the extensive space is directly reflected in the asymmetry of the natural numbers between the finite and the infinite, which in turn is generated by the broken symmetry of the predecessor mapping of the Natural Numbers by the zero, which is the only element that has no predecessor. Also to the intensive space belongs the asymmetry of the Natural Numbers. In this case it is the asymmetry between the finite and the infinite scales, thus between the finitely large and the infinitely small.

An outstanding implication of the asymmetry of the Natural Numbers is shown in the second volume on the continuum, namely the impossibility of the global rearrangement of the Natural Numbers, which in summary means that there is no way out of the extensively or the intensively infinite into the finite. Similarly, the asymmetry in the complexity $\log(N)$ of the first N natural numbers is shown by Eqn. 3.3, i.e., in the inevitability of the existence of short and long names. More generally, the asymmetry is shown by the complexity $\log(\Delta x/u)$ of the continuum according to Eqn. 3.29, which is positive for the extensive space ($\Delta x \geq u$) and negative for the intensive space ($\Delta x < u$). While the limit

$$\lim_{\Delta x \to u} \log\left(\frac{\Delta x}{u}\right) = 0 \tag{5.1}$$

represents the complexity of the finite space, the limits

$$\lim_{\Delta x \to \infty} \log\left(\frac{\Delta x}{u}\right) \to +\infty \tag{5.2}$$

$$\lim_{\Delta x \to 0} \log\left(\frac{\Delta x}{u}\right) \to -\infty \tag{5.3}$$

belong to the extensive or to the intensive infinity of the continuum.

As announced above, let us now turn to the simplest physical systems consisting exclusively of free particles. For this purpose we consider a one-dimensional classical system of exactly two free inertial particles with conservation of mass of each particle, which we interpret in such a way that the particles are neither created nor annihilated. Now we can distinguish two states of motion, contraction and expansion. In the first case, the particles move toward the common center of mass and thus toward each other. The relative distance Δx decreases and with it also the complexity of the two-particle system, while for the expansion the negation is valid:

- Kontraktion: $\frac{d}{dt} \log\left(\frac{\Delta x}{u}\right) < 0$

- Expansion: $\frac{d}{dt}\log\left(\frac{\Delta x}{u}\right) \geq 0$

Mathematically, both states can be realized at will by prescribing appropriate initial conditions. For the physical realization of the contraction there are two possibilities:

- The contraction continues unchanged since unlimited time.
- The state was changed (prepared) at a time of the past, from which follows compellingly, since there are only the two generalized states contraction and expansion for the two free particles, that a change from the expanding to the contracting state has taken place. Otherwise we would be again at the first case.

The first case is ruled out because it presupposes the entry of the two particles from the infinite into the finite, which corresponds to a rearrangement of the order of the extensive infinite space and thus to a global rearrangement of the natural numbers, as described above.

The second case, the inversion of the existing expansion into a contraction is not possible for the system of exactly two free particles, so that the expanding relative state remains as the only possible one and thus the complexity of its relative state can only increase. Overall, then, the complexity – or simply the particle distance – of the unbounded two-particle system is monotonically increasing with time, and any change of relative state is irreversible.

The same is true for an extended one-dimensional classical system of finitely many particles with elastic collisions between neighboring particles. None of the particles, including the two outermost (peripheral) particles, can have entered the system from infinite space into the finite marked by the common center of mass. Thus, the two peripheral particles can only expand relative to each other. And a preparation of their relative state with the aim of reversing their expansion into contraction is not possible with collisions self-referentially brought about by the system from its interior. Rather, each collision of a peripheral particle with its neighbor leads to the increase of the relative velocity between this particle and the other peripheral particle.

Therefore, for a collective consisting of elastically colliding particles, the relative complexity of the two peripheral particles is monotonically increasing and each change of state again is irreversible.

The same is true for three-dimensional collectives of free or elastically colliding particles, since the law of conservation of momentum, which forms the basis for the above argument, is valid along any axis, so that the particles also expand irreversibly along any such axis and thus ultimately in three-dimensional space.

The condition of only finitely many particles is essential because infinitely many can also fill infinite space, so there are no peripheral particles on which to base the simple argument. The finitely many particles make the system nevertheless a system of finite expansion in spite of its basically spatial unboundedness.

Summarizing one can say that the asymmetry of time cannot be derived from the classical equations of motion, but – for free and also for elastically colliding particles – from the asymmetry of the extensive physical space, and that in this respect already *Newton's first axiom*[5] implies the asymmetry of physical processes. For the force-free body moves straight and uniformly and obeys the local symmetry of space and is thus at the same time subject to its global asymmetry. It can move away from an observer – a second body – or approach it. But while the one state of motion, the expansion, is locally possible without the concomitant of a simultaneous contraction, the other, the contraction, is necessarily connected with an expansion in the surrounding area due to the law of conservation of momentum.

---

[5] https://en.wikipedia.org/wiki/Newton's_laws_of_motion

Fundamental to the argument above is that there can be no motion from infinite extensive space into finite space. With finite velocities a movement from the finite into the infinite is just as little possible. So physics takes place in finite space. And in finite space, expansion is the predominant state, or any contraction is not possible without simultaneous expansion. A plate, which falls on the ground, does not collapse, but breaks apart.

5.1.2 Repulsive forces

It is clear that a one-dimensional two-particle system with a repulsive force of unlimited range and otherwise the same conditions as those of the last section expands exclusively, in position space and in this case additionally also in velocity space.

The case of short range, such as that between electrically neutral molecules, is only of interest when the two particles are involved in a larger system with an external environment, so that the conditions of the last section do not apply and the particles can therefore also move towards each other. The limited spatial range has also the limited temporal duration of the interaction (impact) as consequence, so that the impact divides the time into a before and an after with vanishing relative potential in each case. Therefore, according to the law of conservation of energy, the kinetic relative energy of the two particles after the collision is the same as before the collision. However, for the duration of the collision, relative velocity and distance both decrease initially until they both increase at the reversal point, with the increase in velocity ending when the original velocity is reached.

The final effect of the impact, however, does not differ in principle from the interaction-free passage of free particles, whose relative velocity remains unchanged over the whole time of the passage and whose distance also first shortens to a minimum and then lengthens indefinitely, so that the impacting two-particle system with respect to its expansion differs from the system of free particles of the last section in principle only by the changed preconditions which make an initial contraction and thus the impact – instead of the passage – possible at all.

5.1.3 Attractive forces

It is the attractive forces which can reverse the otherwise strictly monotonous expansion of free particles in position space into contraction. For the contraction in the phase space, however, even the attracting natural forces seem to have no meaning, because the decrease of the potential energy (by contraction in the position space) is accompanied by the increase of the kinetic energy (by expansion of the velocities). For this purpose, let us again consider a simple one-dimensional two-particle system with relative distance x, relative velocity v, attractive relative potential $-\gamma x^{-n}$ and *reduced mass*[6] m. Then, according to the law of conservation of energy, Eqn. 5.26, the following holds:

$$\frac{m}{2}v^2 - \gamma x^{-n} = const.$$

In this case, unlike repulsive forces, it is possible to give the value $const. = 0$ to the total energy, so that

$$v^2 \cdot x^n = \frac{2\gamma}{m}$$

holds, from where

$$n\log(|x|) + 2\log(|v|) = \log\left(\frac{2\gamma}{m}\right) = const. \tag{5.4}$$

---

[6] https://en.wikipedia.org/wiki/Reduced_mass

follows. Thus, for this system, if the complexity is defined in terms of a propagation quantity in phase space by the left-hand side of the above equation, then it is constant in time, so that the decrease in position space is compensated by the simultaneous increase in velocity space and vice versa.

Unlike the force-free and the repulsive two-particle system, the two-particle system bound by attractive forces does not show strict monotonicity, but at least simple monotonicity in phase space, while in space and velocity space cyclic changes take place, as it can be seen e.g. in the planetary motion.

Finally, the equation 5.4 sketchily describes also the entropy of a contracting many-body system, as e.g. in the case of the contraction of matter during star formation. So matter contracts in position space with simultaneous expansion of the relative velocities, so that the entropy does not decrease in total in any case.

5.2 Randomness and monotonicity in momentum space

5.2.1 Conservation of momentum and energy

The conservation of momentum and the conservation of energy both follow from *Newtons*[7] decomposition of the many-particle problem into relations between each two particles, which bring the two-particle relative states into the center of consideration. The two theorems are valid for closed physical systems, i.e. just the systems to which also the thesis refers that their entropy does not decrease. Moreover, the connection between energy and entropy is fundamental, for which reason *Clausius*[8] deliberately constructed the name entropy in a word-creative way to accentuate their close connection.

For the above reasons, it is worthwhile to prove the propositions by pointing out the simplicity of the premises made, which are the following (Cf. *Newton's laws of motion*[9], *Newton's law of gravity*[10] and the *Coulomb's law*[11]):

Relative forces and their addition: The dependence of particles among themselves can be decomposed into pairwise relations (forces) between each two particles, which for each particle add up to a total force by simple (vectorial) summation. The force on a particle is thus the (simple) sum of the forces of all other particles, which these exert on that particle.

Force and acceleration: The force $\mathbf{F}$ on a particle results in an acceleration $\mathbf{a}$, which is inversely proportional to the inertial mass m of the particle:

$$m \cdot \mathbf{a} = \mathbf{F}$$

Actio=Reactio: The force between every two particles is symmetric and double-referential, i.e. their sum is zero, and the particles reference each other. From this it follows that the compound of two and consequently also of many particles – in the same way as according to Newton's 1st axiom also each particle for itself alone – does not exert any force on itself (the compound) and thus, as it is shown by the proof of the law of conservation of momentum below, the inner total work done is always equal to zero, from which it is directly deduced that a *perpetuum mobile of the first kind*[12] cannot exist.

---

[7] https://en.wikipedia.org/wiki/Isaac_Newton

[8] https://en.wikipedia.org/wiki/Rudolf_Clausius

[9] https://en.wikipedia.org/wiki/Newton's_laws_of_motion

[10] https://en. wikipedia.org/wiki/Newton's_law_of_universal_gravity

[11] https://en.wikipedia.org/wiki/Coulomb's_law

[12] https://en.wikipedia.org/wiki/Perpetual_motion

Minimization of symmetry breaking: The two-particle force acts in the direction of the symmetry axis, i.e., the line connecting the two particles, and otherwise depends only on their distance, which is symmetrical with respect to the two particles, but with respect to space means that there is no further symmetry breaking of space beyond that which exists anyway as a result of the mere existence of the two particles.

Conservation of momentum follows from the first three postulates above:

Let m_i be the mass of the $i.$ particle and $\mathbf{a}_{ij}$ the acceleration of this particle due to the force exerted by the $j.$ particle. From Actio = Reactio follows:

$$m_i\mathbf{a}_{ij} + m_j\mathbf{a}_{ji} = 0 \tag{5.5}$$

Since further for fixed index i the sum $\sum_{j\neq i}\mathbf{a}_{ij} = \mathbf{a}_i$ is the resulting acceleration of the $i.$ particle due to the action of all other particles, the following applies

$$\sum_{j\neq i} m_i\mathbf{a}_{ij} = m_i\mathbf{a}_i \tag{5.6}$$

and analogously for fixed j $\sum_{i\neq j} m_j\mathbf{a}_{ji} = m_j\mathbf{a}_j$, so that from Eqn. 5.5 with summation over all indices i, j with $i \neq j$

$$0 = \sum_{i\neq j}(m_i\mathbf{a}_{ij}+m_j\mathbf{a}_{ji}) = \sum_i\sum_{j\neq i} m_i\mathbf{a}_{ij}+\sum_j\sum_{i\neq j} m_j\mathbf{a}_{ji} = \sum_i m_i\mathbf{a}_i+\sum_j m_j\mathbf{a}_j = 2\sum_i m_i\mathbf{a}_i \tag{5.7}$$

follows and thus also:

$$\sum_i m_i\mathbf{a}_i = 0 \tag{5.8}$$

In the isolated many-particle system not only the sum of the relative forces between two particles disappears, but as a consequence the sum of all forces in total. It is a simple and well-known fact that from Eqn. 5.8 the law of conservation of momentum follows by integration of the equation. On the other hand, the conservation of momentum is of such fundamental importance that it is worth mentioning. So let v_i be the velocity, $\mathbf{p}_i = m_i\mathbf{v}_i$ be the momentum of the $i.$ particle and $\mathbf{P}$ the total momentum, then it follows from Eqn. 5.8 by integration:

$$\sum_i m_i\mathbf{v}_i = \sum_i \mathbf{p}_i = \mathbf{P} = const. \tag{5.9}$$

One of the essential meanings of the law of conservation of momentum is that it partially decouples the state in momentum space from the state in position space, which is in the same way a feature of the law of conservation of energy, which is now to be proved, for which we need, beyond the first three postulates, also the assumption of the fourth postulate:

Let $\mathbf{r}_i$ be the position vector of the $i.$ particle and $f_{ij} = f_{ji}$ the magnitude of the force of the $j.$ Particle on the $i.$ particle. Then, according to the fourth postulate above:

$$m_i\mathbf{a}_{ij} = -f_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_i - \mathbf{r}_j) \tag{5.10}$$

Summing these equations over the index j and considering 5.6, the equations of motion follow:

$$m_i\mathbf{a}_i = -\sum_j f_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_i - \mathbf{r}_j) \tag{5.11}$$

However, to derive the law of conservation of energy, we stay with Eqn. 5.10. For the longitudinal motion, i.e., the motion of the particle i in the direction of the connecting axis $\mathbf{r}_i - \mathbf{r}_j$, which – and only that – is associated with an acceleration by the other particle, follows:

$$m_i\mathbf{a}_{ij}\mathbf{v}_i = -f_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_i - \mathbf{r}_j)\mathbf{v}_i \tag{5.12}$$

Since because of Actio $=$ Reactio likewise

$$m_j\mathbf{a}_{ji}\mathbf{v}_j = -f_{ij}(|\mathbf{r}_j - \mathbf{r}_i|)(\mathbf{r}_j - \mathbf{r}_i)\mathbf{v}_j \tag{5.13}$$

is valid, it follows by adding the last two equations:

$$m_i\mathbf{a}_{ij}\mathbf{v}_i + m_j\mathbf{a}_{ji}\mathbf{v}_j = -f_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_i - \mathbf{r}_j)(\mathbf{v}_i - \mathbf{v}_j) \tag{5.14}$$

For the left part of the left-hand side of 5.14, because of Eqn. 5.6 we get after summation over the index j:

$$\sum_{j\neq i} m_i\mathbf{a}_{ij}\mathbf{v}_i = m_i\mathbf{a}_i\mathbf{v}_i$$

Analogously, for the right part $m_j\mathbf{a}_{ji}\mathbf{v}_j$ on the left side of the equation 5.14 after summation over i we get the sum $\sum_{i\neq j} m_j\mathbf{a}_{ji}\mathbf{v}_j = m_j\mathbf{a}_j\mathbf{v}_j$. Thus, for both parts of the left-hand side of Eqn. 5.14 together:

$$\sum_{i<j}(m_i\mathbf{a}_{ij}\mathbf{v}_i + m_j\mathbf{a}_{ji}\mathbf{v}_j) = \frac{1}{2}\sum_{i\neq j}(m_i\mathbf{a}_{ij}\mathbf{v}_i + m_j\mathbf{a}_{ji}\mathbf{v}_j) = \frac{1}{2}\left(\sum_i m_i\mathbf{a}_i\mathbf{v}_i + \sum_j m_j\mathbf{a}_j\mathbf{v}_j\right) = \sum_i m_i\mathbf{a}_i\mathbf{v}_i \tag{5.15}$$

Thus it follows from Eqn. 5.14

$$\sum_i m_i\mathbf{a}_i\mathbf{v}_i = \sum_{i<j} -f_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_i - \mathbf{r}_j)(\mathbf{v}_i - \mathbf{v}_j)$$

and finally, if ϕ_{ij} is the primitive of f_{ij}, by integration the law of conservation of energy:

$$\sum_i \frac{m_i}{2}\mathbf{v}_i^2 = -\sum_{i<j}\phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) + \mathit{const.} \tag{5.16}$$

The expression $T = \sum_i \frac{m_i}{2}\mathbf{v}_i^2$ is the kinetic energy of the particle system, while the expressions $\phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|)$ reflect the pairwise potential energies of the system. The law of conservation of energy states that the sum of the kinetic and potential energies is constant in time. The significant thing about this theorem is not just the constancy in time of two summands, but the peculiarity of these summands that one is just a function of velocities and the other is just a function of relative particle distances, which is ultimately the reason for the definition of kinetic energy and potential energy, respectively. With the momentum

$$\mathbf{p}_i = m_i \mathbf{v}_i \tag{5.17}$$

follows the momentum representation of the theorem:

$$\sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i<j} \phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) = const. \tag{5.18}$$

In 5.18 the potential energy depends exclusively on the relative coordinates $\mathbf{r}_i - \mathbf{r}_j$, into which now analogously also the kinetic energy shall be brought. Let be the total mass M, total momentum $\mathbf{P}$, the two-particle mass m_{ij} and the two-particle relative momentum $\mathbf{p}_{ij}$:

$$M := \sum_i m_i \tag{5.19}$$

$$\mathbf{P} := \sum_i \mathbf{p}_i \tag{5.20}$$

$$m_{ij} := \frac{m_i m_j}{M} \tag{5.21}$$

$$\mathbf{p}_{ij} := m_{ij}(\mathbf{v}_i - \mathbf{v}_j) = \frac{m_j}{M}\mathbf{p}_i - \frac{m_i}{M}\mathbf{p}_j \tag{5.22}$$

Then applies

$$\mathbf{p}_{ij} = -\mathbf{p}_{ji} \tag{5.23}$$

and:

$$\begin{aligned}
\sum_{i<j} \frac{\mathbf{p}_{ij}^2}{2m_{ij}} &= \sum_{i<j} \frac{m_{ij}}{2}(\mathbf{v}_i - \mathbf{v}_j)^2 = \frac{1}{2}\sum_{i,j} \frac{m_{ij}}{2}(\mathbf{v}_i^2 - 2\mathbf{v}_i\mathbf{v}_j + \mathbf{v}_j^2) \\
&= \frac{1}{2M}\left(\sum_{i,j} m_j \frac{1}{2} m_i \mathbf{v}_i^2 - \sum_{i,j} \mathbf{p}_i \mathbf{p}_j + \sum_{i,j} m_i \frac{1}{2} m_j \mathbf{v}_j^2\right) \\
&= \frac{T}{2} - \frac{\mathbf{P}^2}{2M} + \frac{T}{2} = T - \frac{\mathbf{P}^2}{2M}
\end{aligned}$$

The kinetic energy according to Eqn. 5.16 f. thus disassembles into its inner and outer part

$$T = \sum_{i<j} \frac{\mathbf{p}_{ij}^2}{2m_{ij}} + \frac{\mathbf{P}^2}{2M} \tag{5.24}$$

from which for the law of conservation of energy also the representation

$$\frac{\mathbf{P}^2}{2M} + \sum_{i<j} \left(\frac{\mathbf{p}_{ij}^2}{2m_{ij}} + \phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \right) = const. \tag{5.25}$$

or

$$\frac{\mathbf{P}^2}{2M} + \sum_{i<j} \left(\frac{m_{ij}}{2} (\mathbf{v}_i - \mathbf{v}_j)^2 + \phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \right) = const. \tag{5.26}$$

follows, so that the total internal energy is the sum of the two-particle relative energies:

$$\frac{\mathbf{p}_{ij}^2}{2m_{ij}} + \phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) = \frac{m_{ij}}{2} (\mathbf{v}_i - \mathbf{v}_j)^2 + \phi_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \tag{5.27}$$

While the law of conservation of momentum makes a statement about the state of the momentum and its changes which is completely decoupled from the state in position space, the law of conservation of energy decouples the state of motion from the state in position space only partially, but shows a second decoupling besides:

- The first decoupling is the decomposition of the constant energy into two parts. The first part, the kinetic energy, depends exclusively on the relative motions of the particles, the second part exclusively on the relative distances in position space.
- The second decoupling is the decomposition of the constant energy into the two-particle relative energies 5.27, which means that the relative state between each two particles i and j evolves in principle like that of a particle in a potential ϕ_{ij}, but this evolution is subject to a continuous perturbation.

Finally, it may be noted that for the conservation of energy the fourth of the postulates listed at the beginning, which is reflected in Eqn. 5.10 is necessary. It means that the relative forces between two particles can be only longitudinal, thus have the direction of the connecting axis. On the other hand, relative transversal forces do not contradict the law of conservation of momentum, as long as their sum vanishes, as required by Actio = Reactio. Two particles at rest, for example, would move accelerated in a circle, but the total momentum would remain unchanged at zero. The angular momentum as well as the kinetic energy would increase steadily. Since for the rotation the size of the position space, represented e.g. by a closed circle line or the angle interval $[0, 2\pi)$, is finite, there could be no function in this space, which would decrease indefinitely to compensate the indefinitely increasing kinetic energy. Thus, the unboundedness of the extensive space is in principle also a condition for the validity of the law of conservation of energy, at least for repulsive forces. The spatial boundedness of physical systems is finally also possible only by attractive forces.

5.2.2 Relations vs. particles in three-dimensional space

Another criticism against *Boltzmanns H-theorem* is that the assumption of statistical independence of particles prior to a collision anticipates the result, as expressed in the article on the H-theorem in the section on *Loschmidt's paradox*[13]:

> *It turns out that this assumption breaks time reversal symmetry in a subtle sense, and therefore begs the question. Once the particles are allowed to collide, their velocity directions and positions in fact do become correlated (however, these correlations are*

---

[13] https://en.wikipedia.org/wiki/H-theorem#Loschmidt's_paradox

> *encoded in an extremely complex manner). This shows that an (ongoing) assumption of independence is not consistent with the underlying particle model.*[5]

I assume that in this critique *particle model* refers to the interaction model given by *Newton*, which leads to the deterministic equations of motion 5.11 and to the two conservation laws. In particular, I would like to examine more closely below the last sentence of the quotation, with its suggestion that the assumption of permanent independence contradicts the model. Again, it is not clear at first what the author means by independence in the context of the model, since explicit independence assumptions are virtually cornerstones of the model (see below). To be sure, these – explicit – assumptions are not statistical independence. But it is easy to show that these independence assumptions give room also for statistical independence, so that the random exchange of momentum and energy among the particles does not contradict the model in any case. To see this, one can first note that the interaction model decomposes the total state of a system into two parts, one of which is the state in position space and the other the state in velocity or momentum space. And the infinitesimal change of one partial state depends on the other partial state. However, besides this symmetry, there is also a striking asymmetry in the relation of the two partial states and their change, which consists in the fact that the change of the position state depends exclusively on the state of the velocities within the system. In fact, the change of the position state of a particle even depends only on its own state of velocity, while the change of its velocity is the cumulative effect of a multitude of individual causes (forces) that arise from other particles within the system, but may also be external to the system. The individual forces on the right-hand side of Eqn. 5.10 are considered independent of each other and become parity when summed to a resulting force on the right side of Eqn. 5.11. In this context, it is now essential that the individual effects $\mathbf{a}_{ij}$ of the individual forces are also independent of each other or are considered as such. They are also added in parity to the sum of effects $\sum_i \mathbf{a}_{ij} = \mathbf{a}_i$ of the individual forces. Thus, the effects in the form of the velocity or momentum changes are, on the one hand, the consequence of their own respective causes and thus depend on them, on the other hand, each effect as part of all effects on a particle is independent of all other effects, so that the effects – analogous to the causes – satisfy their own law of independence and parity accumulation. I.e. this form of independence results, if one abstracts from the compound of cause and effect, the effect, i.e. the change of momentum. The set of infinitesimal momentum changes of a particle defined at a point in time (and their algorithmic processing) is independent of their origin, so it does not contradict the interaction model if there are random momentum changes as well as caused ones.

One must assume that *Boltzmann* at least suspected the statistical character of the microcosm. In this context, the question also arises whether the classical interaction model – with all its premises – has ever been fully formulated in its entirety. One of these premises is the possibility of a free choice of a rigid frame of reference. Such a reference frame consists of at least one reference point, three different directions and one unit of length. The independence of the theory from the choice of the reference frame and, in particular, of the reference point is expressed by the requirement that the equations of motion must be invariant to transformations, e.g., the *Galilei transformation*[14] or the *Lorentz transformation*[15], associated with the change of the reference frame. The requirement of invariance, however, presupposes that the choice in the particular case is possible at all with sufficient accuracy and, if made, is stable. This is at least the basic problem which arises in reality and especially in that of the microcosm, and which is aggravated with increasing demand for precision and smaller scales. In connection with this, the choice of a real reference point is only possible under the (self-referential) reference to matter and its state, i.e. under reference to the broken symmetry of space induced by matter. In other words, a reference point must be a point physically distinguished by matter, which conversely forms the

---

[14] https://en.wikipedia.org/wiki/Galilean_transformation

[15] https://en.wikipedia.org/wiki/Lorentz_transformation

basis for the description of the state of this very matter in space, and it must be this for the duration of the physical process. The change to another reference point requires the readjustment of the system state. Against this it may be objected that the change is unproblematic because the state actually and essentially results from the composition of the relative states $\mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{v}_i - \mathbf{v}_j$ between each two particles i and j. If this is so, however, it must be noted that, in proportion, relative states, which are relations between particles, receive little attention in classical theory except for *Newton's* axioms. First of all, the relations of change (Eqn. 5.10), the forces, belong to the foundation of *Newton's* theory, at least in the interaction model. The summation of these relations, i.e., the transition from Eqn. 5.10 to the condensed equations of motion 5.11, puts the relative states $\mathbf{v}_{ij} := \mathbf{v}_i - \mathbf{v}_j$ and thus the relations between the particles into the background and instead the changes $\mathbf{a}_i$ of the particle states $\mathbf{v}_i$ and thus the particles themselves into the foreground. If instead a particle pair (i, j) is granted independence from third particles, say from a particle k, in addition to the change relation also for its state relation $\mathbf{v}_{ij}$, then the particle states $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ must necessarily smear, just as the vertices of a triangle smear when its side lengths are independently changed and finally reassembled. In the case of completely independent evolution, even for arbitrarily short durations, the laws of momentum and conservation of energy would also surely be violated. I.e., in order to satisfy the theorems, an interaction between particles i and j and the resulting change of state $\mathbf{v}_{ij}$ for a third particle k must be accompanied by the adjustment of states $\mathbf{v}_{ik}$ and $\mathbf{v}_{jk}$, which is a crucial advantage of the reference frame in relation to which the matching is done implicitly and with uniqueness, in that $\mathbf{v}_k$ simply remains unchanged if k itself does not interact.

In a summary, it can be said that the interaction model *Newton's* admits two interpretations, depending on whether one gives priority to the particles or their mutual relations within the framework of the model. The priority of the particles and their unique state is closely connected with the choice of a unique and stable reference frame and, in particular, a reference point in relation to which the definition of the particle in terms of its state – independent of other particles – is initially possible. If this assumption is omitted, the relations between the particles are in the foreground, whereby the sharpness of the particle state is necessarily impaired, as will be shown below.

The laws of conservation of momentum and conservation of energy thereby get a different meaning. They are no longer derived theorems, but postulates for systems which comprise at least three particles. In particular they are postulates for many-particle systems. The evolution of state of a two-particle system, on the other hand, is completely determined by the equations of motion. And even within a many-particle system, two particles are not decorrelated by their interaction, but – relative to a reference frame – the correlation between each of the two collision partners and all the other particles is. For the interacting particles themselves, instead of decorrelation, one must even assume that the interaction takes place, as the quote above expresses it, on the basis of exact relative information.

The importance of energy – not as a derived quantity, but as the starting point of physical theory – has a long tradition in physics, evident as early as *Lagrange formalism*[16] and even more so in *Hamiltonian mechanics*[17].

In this context, reference can also be made to *quantum mechanical entanglement*[18], which is seen in physics only as a quantum mechanical phenomenon that does not exist in classical mechanics:

> *The topic of quantum entanglement is at the heart of the disparity between classical and quantum physics: entanglement is a primary feature of quantum mechanics lacking in classical mechanics.*[8]

---

[16] https://en.wikipedia.org/wiki/Lagrangian_mechanics

[17] https://en.wikipedia.org/wiki/Hamiltonian_mechanics

[18] https://en.wikipedia.org/wiki/Quantum_entanglement

However, the basis of entanglement is the conservation laws and the fact that the physical states of a particle system are relations between the particles, and the particles do not have independent states. Thus, in quantum mechanics, the importance of relations between particles becomes very clear. In fact, however, the relations in classical mechanics are no different if one is willing to give priority to the relations over the particles, which necessarily blurs the particle states.

The importance of the relative states actually already in classical mechanics is first shown by the version 5.26 for the conservation of energy theorem, which states that the two-particle relative energies are additive and their sum coincides with the constant energy. Thus, if one assumes that the relative states are sharp, then so is the energy and so are other state variables and averages, such as the velocity variance σ_v^2, for which Eq. 3.44 holds in the same way:

$$\sigma_v^2 = \frac{1}{N}\sum_i \left(\mathbf{v}_i - \frac{1}{N}\sum_j \mathbf{v}_j\right)^2 = \frac{1}{N^2}\sum_{i<j}(\mathbf{v}_i - \mathbf{v}_j)^2$$

In the following it shall be made clear that the – correlated – interaction between two particles each is accompanied by the random momentum and energy exchange of each of these particles with all other particles, at least if one describes the state of the system and its development in relation to a reference point common for all particles, but at the same time assumes that the change of state takes place only on the basis of the relative states, the relative equations of motion 5.10 and the two conservation laws. Thus, we assume that we can compute the changes in relative states based on a fictitious frame of reference and use the results – then assuming that the system has no other information besides this relative information – to check the uniqueness of the change in state.

Accordingly, we assume that the particle i undergoes a change of momentum $\Delta\mathbf{p}_i$ as a result of the interaction with the particle j and the particle j undergoes the change of momentum

$$\Delta\mathbf{p}_j = -\Delta\mathbf{p}_i \tag{5.28}$$

in relation to a reference frame. For the relative momentum $\mathbf{p}_{ik}$ and $\mathbf{p}_{jk}$ with respect to a third particle k holds:

$$\Delta\mathbf{p}_{ik} = \left(\frac{m_k}{M}(\mathbf{p}_i + \Delta\mathbf{p}_i) - \frac{m_i}{M}\mathbf{p}_k\right) - \left(\frac{m_k}{M}\mathbf{p}_i - \frac{m_i}{M}\mathbf{p}_k\right) = \frac{m_k}{M}\Delta\mathbf{p}_i$$

So it follows from Eqn. 5.28 the conservation of momentum for the two-particle momenta $\mathbf{p}_{ik}$ and $\mathbf{p}_{jk}$ when the particles i and j interact:

$$\Delta\mathbf{p}_{jk} = -\Delta\mathbf{p}_{ik} = -\frac{m_k}{M}\Delta\mathbf{p}_i \tag{5.29}$$

For the interacting particles i and j themselves holds

$$\Delta\mathbf{p}_{ij} = \left(\frac{m_j}{M}(\mathbf{p}_i + \Delta\mathbf{p}_i) - \frac{m_i}{M}(\mathbf{p}_j + \Delta\mathbf{p}_j)\right) - \left(\frac{m_j}{M}\mathbf{p}_i - \frac{m_i}{M}\mathbf{p}_j\right)$$

and thus:

$$\Delta\mathbf{p}_{ij} = \frac{m_i + m_j}{M}\Delta\mathbf{p}_i \tag{5.30}$$

From eqs. 5.28, 5.29 and 5.30 it follows:

$$\sum_{k\neq i}\Delta\mathbf{p}_{ik}=\Delta\mathbf{p}_i=-\Delta\mathbf{p}_j=-\sum_{k\neq j}\Delta\mathbf{p}_{jk} \tag{5.31}$$

For the change ΔT_{ik} of the kinetic relative energy of the particles i and k

$$\Delta T_{ik}:=\frac{(\mathbf{p}_{ik}+\Delta\mathbf{p}_{ik})^2}{2m_{ik}}-\frac{\mathbf{p}_{ik}^2}{2m_{ik}}=\frac{1}{m_i}\Delta\mathbf{p}_i\mathbf{p}_{ik}+\frac{m_k}{M}\frac{\Delta\mathbf{p}_i^2}{2m_i}$$

is valid and thus also:

$$\Delta T_{ik}=\frac{m_k}{M}\left(\Delta\mathbf{p}_i(\mathbf{v}_i-\mathbf{v}_k)+\frac{\Delta\mathbf{p}_i^2}{2m_i}\right) \tag{5.32}$$

Analogously, ΔT_{jk} is obtained, so that together with Eqn. 5.28

$$\Delta T_{jk}=\frac{m_k}{M}\left(\Delta\mathbf{p}_i(\mathbf{v}_k-\mathbf{v}_j)+\frac{\Delta\mathbf{p}_i^2}{2m_j}\right) \tag{5.33}$$

and finally

$$\Delta T_{ik}+\Delta T_{jk}=\frac{m_k}{M}\left(\Delta\mathbf{p}_i(\mathbf{v}_i-\mathbf{v}_j)+\frac{\Delta\mathbf{p}_i^2}{2m_im_j}(m_i+m_j)\right) \tag{5.34}$$

follows.

For the interacting particles i and j themselves, because of Eqn. 5.28

$$\Delta T_{ij}:=\frac{(\mathbf{p}_{ij}+\Delta\mathbf{p}_{ij})^2}{2m_{ij}}-\frac{\mathbf{p}_{ij}^2}{2m_{ij}}=\frac{m_i+m_j}{m_{ij}}\mathbf{p}_{ij}\Delta\mathbf{p}_i+\frac{\Delta\mathbf{p}_i^2}{2m_{ij}}\left(\frac{m_i+m_j}{M}\right)^2$$

holds and therefore also:

$$\Delta T_{ji}=\Delta T_{ij}=\frac{m_i+m_j}{M}\left(\Delta\mathbf{p}_i(\mathbf{v}_i-\mathbf{v}_j)+\frac{\Delta\mathbf{p}_i^2}{2m_im_j}(m_i+m_j)\right) \tag{5.35}$$

This Eqn. together with Eqn. 5.34 and $\Delta T_{ii}:=0$ results in:

$$\sum_k(\Delta T_{ik}+\Delta T_{jk})=\Delta\mathbf{p}_i(\mathbf{v}_i-\mathbf{v}_j)+\frac{\Delta\mathbf{p}_i^2}{2m_im_j}(m_i+m_j) \tag{5.36}$$

Now this sum is identical with the change of the kinetic energy of the two interacting particles i and j in relation to the common reference point, because the following holds true

$$\Delta T_i+\Delta T_j=\frac{(\mathbf{p}_i+\Delta\mathbf{p}_i)^2}{2m_i}-\frac{\mathbf{p}_i^2}{2m_i}+\frac{(\mathbf{p}_j+\Delta\mathbf{p}_j)^2}{2m_j}-\frac{\mathbf{p}_j^2}{2m_j}=\frac{\mathbf{p}_i\Delta\mathbf{p}_i}{m_i}-\frac{\mathbf{p}_j\Delta\mathbf{p}_i}{m_j}+\frac{\Delta\mathbf{p}_i^2}{2m_i}+\frac{\Delta\mathbf{p}_j^2}{2m_j}$$

from where we get:

$$\Delta T_i + \Delta T_j = \Delta \mathbf{p}_i(\mathbf{v}_i - \mathbf{v}_j) + \frac{\Delta \mathbf{p}_i^2}{2m_i m_j}(m_i + m_j) \tag{5.37}$$

In summary one can state: The interaction of two particles i and j with their conservation of momentum according to Eqn. 5.28 and energy change according to Eqn. 5.37 in relation to a reference point common to all particles presents itself – in a description omitting the reference point and that instead is based on the relative information – in such a way that the interaction of particles i and j also involves all other particles $k \neq i, j$. The involvement is such that the relative momenta $\mathbf{p}_{ik}$ and $\mathbf{p}_{jk}$ behave in their relation to each other like the interacting particles i and j. Thus, in summary, according to the equations 5.28 to 5.37:

$$\Delta \mathbf{p}_i + \Delta \mathbf{p}_j = 0$$

$$\Delta \mathbf{p}_{ik} + \Delta \mathbf{p}_{jk} = 0$$

$$\Delta T_i + \Delta T_j = \Delta \mathbf{p}_i(\mathbf{v}_i - \mathbf{v}_j) + \frac{\Delta \mathbf{p}_i^2}{2m_i m_j}(m_i + m_j)$$

$$\Delta T_{ij} = \frac{m_i + m_j}{M}(\Delta T_i + \Delta T_j)$$

$$\Delta T_{ik} + \Delta T_{jk} = \frac{m_k}{M}(\Delta T_i + \Delta T_j)$$

However, besides this agreement, there is also an essential difference which consists in the fact that the particle interaction is subject to the relative potential ϕ_{ij} and thus to the correlative accompaniment by the equations of motion, which, however, does not apply to the implicit interaction between the relative momenta $\mathbf{p}_{ik}$ and $\mathbf{p}_{jk}$ in three-dimensional space, whose change $\Delta \mathbf{p}_{ik}$ and $\Delta \mathbf{p}_{jk}$ with six unknowns is therefore subject only to the conservation of momentum $\Delta \mathbf{p}_{ik} + \Delta \mathbf{p}_{jk} = 0$ and the given energy change $\Delta T_{ik} + \Delta T_{jk}$ with a total of only four equations, so that these equations only incompletely specify the momentum change, opening a random window for the relative motion. This, in turn, is not necessarily equivalent to the relative information losing its sharpness, because the outcome of a random process, such as that of a throwing experiment, is very much sharp. On the other hand, there is no unambiguous reconstruction of the particle states from the changed relative states, because, conversely, unambiguous changes of the relative states would be derived from the unambiguousness of the particle states and their changes.

Altogether, the assumption that the relative information is the primary information, or at any rate the information on the basis of which the interaction process proceeds for the participants and all those not directly involved, leads to the result that the process definition in the form of the equations 5.10 is incomplete and therefore the change in the relative states between the colliding particles on the one side and the other particles on the other side is not unambiguously determined and, as a consequence, the particle states in relation to a reference frame become blurred.

Before proceeding further, the following statement by *Edwin Thompson Jaynes* should be quoted:

> *Nevertheless, we still see attempts to "explain irreversibility" by searching for some entropy function that is supposed to be a property of microstate, making the second law a theorem of dynamics, a consequence of the equations of motion. Such attempts, dating back to Boltzmann's paper of 1866, have never succeeded and never ceased. But they are quite unnecessary; for the second law that Clausius gave us was not a statement about any property of microstates. The difference in dS on mixing of like and unlike gases can seem paradoxical only to one, who supposes erroneously, that entropy is a property of the microstate.*[2]

As so far in this chapter, I understand by the term irreversibility in the evolution of a system the fact that there is a monotonic state function, and that because of this monotonicity the system cannot, once the function values have changed in the course of time, return to a state prior to that change:

Position space: In the case of the position space, this function is related to the expansion in the **extensive** position space. This expansion is a direct consequence of the equations of motion, and it does not change the correlation of the particles.

Momentum space: In contrast, the result of this section is that, first, the monotonicity in momentum space is identical with the monotonic decrease of the momentum correlations and, second, this – decrease – although not a direct consequence of the equations of motion, does arise in their context. Thirdly, the decrease of the momentum correlations is an expansion in the **intensive** momentum space, since – at least in isolated systems – the standard deviation of the momentum correlates with the energy and thus the extensive quantity remains unchanged.

Chapter 6

Conclusions from the entropy of the ideal gas

6.1 The equilibrium entropy of the ideal gas

The ideal gas is the mental concept of a physical system of quasi-free particles, in which the potential energy of the mechanical interaction between the particles is neglected, which does not mean that collisions between the particles or with a bounding wall are also neglected. Only the time of the interaction process is set to zero. Thus, we are dealing with hard, completely elastic collisions which take place in such a short time that the potential energy, without which collisions are basically impossible, is nevertheless negligible in the energy balance locally and globally, so that the energy, locally as well as globally and at any time, is exclusively kinetic energy.

The ideal gas is thus the abstraction of a many-particle system with restriction to the kinematics or the kinetic relations between the particles, which can be interpreted only statistically because of the large number of particles. However, the statistics is assigned not only an interpretative role, but rather the significant role. And it is to be assumed that entropy is exactly that, a descriptive, statistical measure namely of the kinematic state under omission of the question, how it came to this state and likewise of the question, what can still become from this state by consideration of the mechanical interactions. The second law of thermodynamics in its general form, according to which the entropy of an isolated system does not decrease, independent of the concrete form of the internal mechanical interactions, indicates in this context that there must be something beyond mechanics, which is common to all physical systems. The past chapters indicate that it is essentially the structure of space or the kinematic relations between space and matter, i.e. the distribution of matter and that of its motion in space. Conversely, it is matter and its motion that gives space its hatching and structure in the first place and thus creates a reference system for itself. This mutual reference of space and matter resembles the double reference between the two parts of the double stream[1] described by the Schrödinger equation.

Apparently, then, the kinematics is the exclusive focus of entropy, so that the ideal gas is presumably not accidentally the starting point of the mental development to entropy, but an abstraction which captures the essence of entropy in general.

For the *equilibrium entropy*[2] S of the monatomic ideal gas holds:

$$S(N,V,T) - S(N,V_R,T_R) = N k_B \log\left(\frac{V}{V_R}\right) + \frac{3}{2} N k_B \log\left(\frac{T}{T_R}\right) \tag{6.1}$$

---

[1] Cf. the section 2.3.2 on the Schrödinger equation

[2] https://en.wikipedia.org/wiki/Ideal_gas#Entropy

This equation gives the entropy $S = S(N, V, T)$ of a macroscopic state in thermodynamic equilibrium with given particle number N, volume V, and temperature T relative to the entropy $S = S(N, V_R, T_R)$ of a reference state N, V_R, T_R.

6.2 Entropy of the disequilibrium

In the following, we will derive from the entropy of equilibrium a relation for systems which are not in equilibrium. The only condition is that the entropy $S = S(N, V, T)$, like the number of particles and the volume, is itself a linearly extensive state variable.[3] Namely, because of extensivity, for $0 \leq p$

$$pS(N, V, T) = S(pN, pV, T) \tag{6.2}$$

holds, which is similar to the theorem of intersecting lines for the three edge lengths N, V, S and pN, pV, pS. Moreover, because of Eqn. 6.1 we get

$$S(pN, V, T) - S(pN, pV, T) = pNk_B \log\left(\frac{V}{pV}\right) = -Nk_Bp\log(p)$$

thus due to Eqn. 6.2 also:

$$S(pN, V, T) = pS(N, V, T) - Nk_Bp\log(p) \tag{6.3}$$

If now the whole system of the gas is decomposed into cells C_i of equal size with equal volume V_R, individual particle numbers p_iN and temperatures T_i and it is assumed that in each cell thermodynamic equilibrium prevails for itself, i.e. for the entropy $S(p_iN, V_R, T_i)$ of the cell C_i Eqn. 6.3 holds, viz.

$$S(p_iN, V_R, T_i) = p_iS(N, V_R, T_i) - Nk_Bp_i\log(p_i)$$

then because of the extensivity of the entropy, because therefore also the entropy S_{NE} of the non-equilibrium is equal to the sum of its partial entropies, we get:

$$S_{NE} = \sum_i S(p_iN, V_R, T_i) = -Nk_B\sum p_i\log(p_i) + \sum_i p_iS(N, V_R, T_i) \tag{6.4}$$

Now, for the sum of the right-hand expressions in 6.4 with respect to a common reference temperature T_R, e.g., an average temperature or a minimum temperature, it follows using Eqn. 6.1:

$$\begin{aligned}\sum_i p_iS(N, V_R, T_i) &= S(N, V_R, T_R) + \sum_i p_iS(N, V_R, T_i) - S(N, V_R, T_R)\\ &= S(N, V_R, T_R) + \frac{3}{2}Nk_B\sum_i p_i\log\left(\frac{T_i}{T_R}\right)\end{aligned}$$

From Eqn. 6.4 thus becomes:

---

[3] Cf. the section 3.5.2 on the problem of extensivity of Shannon entropy

$$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i p_i \log(p_i) + \frac{3}{2} Nk_B \sum_i p_i \log\left(\frac{T_i}{T_R}\right) \tag{6.5}$$

In this equation, the index i can be replaced by a double index ij, in which the first – global – index i represents physically homogeneous cell neighborhoods, each consisting of a number of neighboring cells with the same physical properties, while the second – local – index j distinguishes the cells within each neighborhood. From Eqn. 6.5 then follows:

$$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i \sum_j p_{ij} \log(p_{ij}) + \frac{3}{2} Nk_B \sum_i \sum_j p_{ij} \log\left(\frac{T_{ij}}{T_R}\right) \tag{6.6}$$

For the entropy of a neighborhood with m_i cells C_{ij}, after assuming equal weights $p := p_{ij}$ and equal temperatures T_{ij}, if $p_i := m_i p, V_i := m_i V_R$, and $T_i := T_{ij}$, the following two equations hold:

$$\sum_j p_{ij} \log(p_{ij}) = m_i p \log(p) = m_i p \log(m_i p) - m_i p \log\left(\frac{m_i V_R}{V_R}\right) = p_i \log(p_i) - p_i \log\left(\frac{V_i}{V_R}\right)$$

$$\sum_j p_{ij} \log\left(\frac{T_{ij}}{T_R}\right) = p_i \log\left(\frac{T_i}{T_R}\right)$$

From Eqn. 6.6 thus becomes

$$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i p_i \log(p_i) + Nk_B \sum_i p_i \log\left(\frac{V_i}{V_R}\right) + \frac{3}{2} Nk_B \sum_i p_i \log\left(\frac{T_i}{T_R}\right) \tag{6.7}$$

or with $N_i := p_i N$:

$$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i p_i \log(p_i) + \sum_i \left[N_i k_B \log\left(\frac{V_i}{V_R}\right) + \frac{3}{2} N_i k_B \log\left(\frac{T_i}{T_R}\right)\right] \tag{6.8}$$

And finally, due to Eqn. 6.1:

$$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i p_i \log(p_i) + \sum_i (S(N_i, V_i, T_i) - S(N_i, V_R, T_R)) \tag{6.9}$$

Instead of, as just done, in Eqn. 6.5, the cells of homogeneous neighborhoods, i.e., cells with the same physical properties, can be grouped together, conversely, each cell C_i can also be decomposed into several cells C_{ij} not necessarily of the same sizes V_{ij}. Because uniform distribution is assumed within C_i, the relative particle numbers p_{ij} of cells C_{ij} in C_i are $p_{ij} = V_{ij}/V_R$. Then, because of $\sum_j p_{ij} = 1$, the following holds further:

$$\begin{aligned} -p_i \log(p_i) &= -\sum_j p_i p_{ij} \log(p_i p_{ij}) + \sum_j p_i p_{ij} \log(p_{ij}) \\ &= -\sum_j p_i p_{ij} \log(p_i p_{ij}) + \sum_j p_i p_{ij} \log\left(\frac{V_{ij}}{V_R}\right) \end{aligned}$$

One could now, at will, for each of the cells C_i in Eqn. 6.5, make corresponding partitionings into subcells C_{ij}, substitute the above result into Eq. 6.5 and would then obtain the following result

$$\begin{aligned}S_{NE} &= S(N,V_R,T_R) - Nk_B\sum_{ij} p_i p_{ij}\log(p_i p_{ij}) + Nk_B\sum_{ij} p_i p_{ij}\log\left(\frac{V_{ij}}{V_R}\right)\\ &+ \frac{3}{2}Nk_B\sum_i p_i\left[\left(\sum_j p_{ij}\right)\log\left(\frac{T_i}{T_R}\right)\right]\end{aligned}$$

from where with $T_{ij} = T_i$

$$\begin{aligned}S_{NE} &= S(N,V_R,T_R) - Nk_B\sum_{ij} p_i p_{ij}\log(p_i p_{ij}) + Nk_B\sum_{ij} p_i p_{ij}\log\left(\frac{V_{ij}}{V_R}\right)\\ &+ \frac{3}{2}Nk_B\sum_{ij} p_i p_{ij}\log\left(\frac{T_{ij}}{T_R}\right)\end{aligned}$$

follows, so again Eqn. 6.7, which means that in Eqn. 6.7 to 6.9 the ratio V_i/V_R can be any, i.e. greater or less than 1.

Equation 6.9 naturally continues what is already contained in equation 6.1, namely that in the context of entropy only relative values are of importance. In Eqn. 6.1 these are the ratios V/V_R and T/T_R, respectively the logarithmic differences $\log(V) - \log(V_R)$ and $\log(T) - \log(T_R)$. In Eqn. 6.9, it is the entropy itself with the reference entropies $S(N,V_R,T_R)$ and $S(N_i,V_R,T_R)$ and the differences $S_{NE} - S(N,V_R,T_R)$ and $S(N_i,V_i,T_i) - S(N_i,V_R,T_R)$, respectively. Taking this into account, according to this equation, the non-equilibrium entropy is equal to the sum of local equilibrium entropies plus the globalization complexity[4].

In any case, the equilibrium entropy $S(N,V_R,T_R)$, which serves as a reference value, is obviously of outstanding importance, i.e. the entropy of a system whose N particles are gathered in the small volume V_R at a small temperature T_R, both of which do not have to be small at all in physical and mathematical respect, but for the purpose of a better view. For then the reference entropy $S(N,V_R,T_R)$, as incidentally also each of the entropies $S(N_i,V_R,T_R)$, is the entropy of a concentrated particle collective, contracted both in position space and in the momentum space. In contrast, the differences $S(N_i,V_i,T_i) - S(N_i,V_R,T_R)$ are the local entropy increases due to local propagation, while the globalization complexity $-Nk_B\sum_i p_i\log(p_i)$ can only be assigned to the global system, which is also shown by the fact that while the absolute particle number Np_i in cell C_i is defined independently of the other cells, the relative particle number p_i in $\log(p_i)$ is not.

Special cases:

- If the system consists of only one cell with volume $V = V_R$ and temperature $T = T_R$, then 6.7 becomes:

  $$S_{NE} = S(N,V,T)$$

  By construction, the system is then in equilibrium and therefore the entropy is also identical to the equilibrium entropy.

---

[4] Cf. definition in section 3.3.2

- If the system is again at equilibrium, but partitioned into M cells with sizes $V_i = V_R$ and temperatures $T_i = T_R$, then 6.7 reduces to:

  $$S_{NE} = S(N, V_R, T_R) - Nk_B \sum_i p_i \log(p_i) = S(N, V_R, T_R) + Nk_B \log(M)$$

  S_{NE} in this case is again itself an equilibrium entropy. There are no local expansions, only the global expansion with complexity $Nk_B \log(M)$.

- In the case where the distribution of particles among cells is constant in time, the time differential of the total entropy is according to Eqn. 6.7:

  $$dS_{NE} = \frac{3}{2} Nk_B \sum_i p_i \frac{dT_i}{T_i} = \sum_i \frac{d\left(\frac{3}{2} p_i N k_B T_i\right)}{T_i} = \sum_i \frac{\delta Q_i}{T_i} \tag{6.10}$$

  Here we assume that we are dealing with a monatomic ideal gas with *internal energy* $E_i = 3/2(p_i N k_B T_i)$ of cell C_i, whose change dE_i we identify with the net heat flux δQ_i from section 2.2.

Using the abbreviations $v_i := V_i/V_R$, $\varrho_i := p_i/v_i$ (see below), and $t_i := T_i/T_R$, Eqn. 6.7 can also be written as follows:

$$S_{NE} = S(N, V_R, T_R) + Nk_B \left(- \sum_i p_i \log(p_i) + \sum_i p_i \log(v_i) + \frac{3}{2} \sum_i p_i \log(t_i) \right) \tag{6.11}$$

Chapter 7

Density entropy and U-entropy

7.1 Definition and normalization

In Eqn. 6.7, the two middle summands correspond to the components of the density entropy defined in Eqn. 3.35, i.e., the Shannon entropy and the complexity of the continuous space, respectively. Correspondingly, we also define here by omitting the particle number N and the Boltzmann constant k_B:

$$S_D = -\sum_{i=1}^{n} p_i \log(p_i) + \sum_i p_i \log\left(\frac{V_i}{V_R}\right) \tag{7.1}$$

And as in the estimation 3.36, using Jensen's inequality 3.30, if $V := \sum V_i$ is the total volume of the system, it follows equally:

$$S_D \leq \log\left(\frac{V}{V_R}\right) \tag{7.2}$$

If we identify the reference volume V_R with the total volume V and set

$$y_i := p_i$$

$$x_i := \frac{V_i}{V}$$

then Eqn. 7.1 – again together with Jensen's inequality – becomes:

$$S_D(1) = -\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{x_i}\right) \leq 0 \tag{7.3}$$

With definition of the relative density[1]

$$\rho_i := \frac{y_i}{x_i}$$

follows:

$$S_D(1) = -\sum_{i=1}^{n} y_i \log\,(\rho_i) = -\sum_{i=1}^{n} x_i \rho_i \log\,(\rho_i) \leq 0 \tag{7.4}$$

---

[1] Distinction from absolute density ϱ is possible only in section 7.3 in the context of partitioning hierarchy

This global entropy is never positive and vanishes exactly when all densities ρ_i are equal. The transition from the sum to the integral leads again to the differential entropy[2]

$$S_D(1) = -\int \varrho(\mathbf{r}) \log\left(\varrho(\mathbf{r})\right) d^3r \leq 0 \tag{7.5}$$

where now we have to integrate over a region with volume 1, and instead of the relative density ρ necessarily the absolute density ϱ defined in section 7.3 must appear in the integral.

By normalizing the volume to 1 and not taking into account the absolute number of particles, the global density entropy, like the Shannon entropy[3], becomes a self-referential description of quantity, since it lacks reference to an other, the external world, and instead reflects exclusively internal relations.

7.2 Change to U-entropy

The definition of the global density entropy according to Eqn. 7.1 places the uniform distribution in the center of consideration, insofar as the entropy of an disequilibrium is always smaller than the upper bound set by the equilibrium. Increasing disequilibrium is thus subject to negative monotonicity, and equating the reference volume V_R with the total volume V in the previous section makes the entropy of equilibrium a zero line, in relation to which the entropy of a disequilibrium as is given by Eqn. 7.3 is always negative. Now, positive monotonicity is, first, more pleasant and, second, the disequilibrium is more interesting than the equilibrium, so it makes sense to associate positive monotonicity with increasing disequilibrium. Accordingly, we define the disequilibrium density U_D entropy (with the U as in *un*):

$$U_D = -S_D = \sum_i p_i \log(p_i) - \sum_i p_i \log\left(\frac{V_i}{V_R}\right) \tag{7.6}$$

With $V_R = V$, the entropy of equilibrium again sets a zero line, but relative to which the entropy of any disequilibrium is now positive. From eqs. 7.3 to 7.5 becomes:

$$U_D(1) = \sum_i y_i \log\left(\frac{y_i}{x_i}\right) \geq 0 \tag{7.7}$$

$$U_D(1) = \sum_i y_i \log\left(\rho_i\right) = \sum_i x_i \rho_i \log\left(\rho_i\right) \geq 0 \tag{7.8}$$

$$U_D(1) = \int \varrho(\mathbf{r}) \log\left(\varrho(\mathbf{r})\right) d^3r \geq 0 \tag{7.9}$$

where again we have to integrate over a region with volume 1.

U-entropy, unlike density entropy, is a contraction measure that measures the magnitude of the mean deviation of density from the mean 1.

---

[2] Cf. section 3.4.4

[3] Cf. the section 3.5.2 on the self-reference of Shannon entropy

7.3 Monotonicity and recursive partitioning

In Eqn. 7.7, x_i is the relative size of a cell C_i in position space and y_i is the relative particle share of this cell in relation to the entire system. Now, if this cell is further partitioned into subcells C_{ij} with – in respect to cell C_i – relative cell sizes x_{ij} and particle shares y_{ij}, so that $\sum_j x_{ij} = \sum_j y_{ij} = 1$, then the shares of cells C_{ij} relative to the whole system are equal to $x_i \cdot x_{ij}$ and $y_i \cdot y_{ij}$. Therefore, for the cumulative share of the entropy of all subcells C_{ij} of cell C_i relative to the whole system holds:

$$\sum_j y_i y_{ij} \log\left(\frac{y_i y_{ij}}{x_i x_{ij}}\right) = y_i \log\left(\frac{y_i}{x_i}\right) + y_i \sum_j y_{ij} \log\left(\frac{y_{ij}}{x_{ij}}\right) \tag{7.10}$$

Thus, the partitioning of cell C_i and its replacement by its subcells C_{ij} implies the global entropy increase

$$y_i \sum_j y_{ij} \log\left(\frac{y_{ij}}{x_{ij}}\right)$$

which is formally equal to the entropy itself according to Eqn. 7.7 except for the factor y_i. If all cells C_i are partitioned in this way into subcells C_{ij} and the partitioning is finally continued recursively, we obtain analogous to the Shannon entropy in Eqn. 3.13 to 3.15 for the U-entropy:

$$\begin{aligned}
U_D(1) &= \sum_i y_i \left(\log\left(\frac{y_i}{x_i}\right) + \sum_j y_{ij}\left(\log\left(\frac{y_{ij}}{x_{ij}}\right) + \sum_k y_{ijk}\left(\log\left(\frac{y_{ijk}}{x_{ijk}}\right) + \ldots\right)\right)\right) && (7.11)\\
&= \sum_i y_i \log\left(\frac{y_i}{x_i}\right) \sum_{ij} y_i y_{ij} \log\left(\frac{y_i y_{ij}}{x_i x_{ij}}\right) \sum_{ijk} y_i y_{ij} y_{ijk} \log\left(\frac{y_i y_{ij} y_{ijk}}{x_i x_{ij} x_{ijk}}\right) + \cdots && (7.12)\\
&= \sum_{ijk\cdots} y_i y_{ij} y_{ijk} \cdots \log\left(\frac{y_i y_{ij} y_{ijk}}{x_i x_{ij} x_{ijk}} \cdots\right) && (7.13)
\end{aligned}$$

As for the Shannon entropy, we give the levels of the resulting hierarchy the designation *scale* and number them down the scale. The undivided whole system thus receives the scale 0, its division described by the index i above, i.e. the set of cells C_i, the scale 1, the set of cells C_{ij} the scale 2, and so on. So, for simplicity, we express smaller scales – contradictorily – by larger scale values n.

With the relative densities

$$\rho_i := \frac{y_i}{x_i}, \rho_{ij} := \frac{y_{ij}}{x_{ij}}, \cdots \tag{7.14}$$

and the absolute densities

$$\varrho_i := \rho_i, \varrho_{ij} := \rho_i \rho_{ij}, \cdots \tag{7.15}$$

is obtained for the first and the third of the above equations:

$$\begin{aligned}
U_D(1) &= \sum_i x_i \left(\rho_i \log(\rho_i) + \sum_j x_{ij}\left(\rho_{ij} \log(\rho_{ij}) + \sum_k x_{ijk}\left(\rho_{ijk} \log(\rho_{ijk}) + \ldots\right)\right)\right) && (7.16)\\
&= \sum_{ijk\cdots} x_i x_{ij} x_{ijk} \cdots \rho_i \rho_{ij} \rho_{ijk} \cdots \log(\rho_i \rho_{ij} \rho_{ijk} \cdots) && (7.17)\\
&= \sum_{ijk\cdots} x_i x_{ij} x_{ijk} \cdots \varrho_{ijk\cdots} \log(\varrho_{ijk\cdots}) && (7.18)
\end{aligned}$$

Thus, U-entropy behaves like Shannon entropy under recursive partitioning. In particular, it is true for both that deep and flat entropy are the same, as shown by the above equations 7.16 and 7.18, and as already expressed for Shannon entropy in section 3.3.3. Therefore, to look at the change characteristic of the entropy downscale, i.e., to determine what happens when the global resolution is increased again by cell division, it is sufficient, as in Eqn. 7.10 above, to assume a flat partitioning representing a complete partitioning hierarchy up to and including a scale n. The global entropy increase $\Delta U_D(1,n)$ from a scale $n-1$ to scale n is then according to that equation:

$$\Delta U_D(1,n) = \sum_i y_i \left(\sum_j y_{ij} \log \left(\frac{y_{ij}}{x_{ij}} \right) \right) \geq 0 \tag{7.19}$$

Therefore, it is the mean of the local entropies $\sum y_{ij} \log(y_{ij}/x_{ij})$ and thus, in any case, also ≥ 0, so that the U-entropy $U_D(1,n)$ is a monotonically increasing function of scale n:

$$U_D(1,0) = 0$$

$$U_D(1,n) = U_D(1,n-1) + \Delta U_D(n) \geq U_D(1,n-1) \tag{7.20}$$

7.4 Local non-negative entropy

According to 7.20, the global U-entropy increases monotonically as a function of scale, and even the division of each cell C_i has a non-negative contribution U_i because – except for a positive factor – according to Eqn. 7.19, it too is of the form:

$$U_i = \sum_j y_{ij} \log \left(\frac{y_{ij}}{x_{ij}} \right) \geq 0 \tag{7.21}$$

However, the summands of this sum, the partial entropies $U_{ij} := y_{ij} \log(y_{ij}/x_{ij})$, are not all ≥ 0. On the contrary, it holds

$$U_i > 0 \Leftrightarrow \exists j : U_{ij} < 0 \tag{7.22}$$

which is because of $U_{ij} < 0 \Leftrightarrow \log(\rho_{ij}) < 0 \Leftrightarrow \rho_{ij} < 1$ and with $\rho_{ij} = y_{ij}/x_{ij}$ equivalent to:

$$U_i > 0 \Leftrightarrow \exists j : \rho_{ij} < 1 \tag{7.23}$$

The proof of this is as follows: By Jensen's inequality 3.30 f., U_i is zero exactly if all densities y_{ij}/x_{ij} are equal, which in turn is the case exactly if they are all equal to 1. Indeed, let ρ be the same value for all densities, then:

$$1 = \sum_j x_{ij} \frac{y_{ij}}{x_{ij}} = \sum_j x_{ij} \rho = \rho$$

Thus, the implication of Jensen's inequality is also the equivalence

$$U_i = 0 \Leftrightarrow \forall j : \rho_{ij} = 1$$

and thus also its contraposition:

$$U_i > 0 \Leftrightarrow \exists j : \rho_{ij} \neq 1$$

This finally is equivalent to the assertion 7.23 if also the following relation holds:

$$\exists j : \rho_{ij} \neq 1 \Leftrightarrow \exists j : \rho_{ij} < 1$$

Of this relation only the implication

$$\exists j : \rho_{ij} \neq 1 \Rightarrow \exists j : \rho_{ij} < 1$$

or its contraposition

$$\forall j : \rho_{ij} \geq 1 \Rightarrow \forall j : \rho_{ij} = 1$$

is to prove: To do this, we assume that for a density $\rho_{ik} > 1$ holds. But then, together with the implication premise, the contradiction

$$1 = \sum_j x_{ij} \frac{y_{ij}}{x_{ij}} = \sum_{j \neq k} x_{ij}\rho_{ij} + x_{ik}\rho_{ik} > \sum_{j \neq k} x_{ij} + x_{ik} = 1$$

follows.

Summarizing up to this point, 7.22 holds with the statement that a growth

$$U_D(1, n+1) > U_D(1, n)$$

of the global entropy[4] is necessarily associated with the emergence of locally negative entropies.

Now one can adapt the definition of entropy according to Eqn. 7.7 by a modification so that the partial entropy U_i of a cell remains unchanged in the sum, but even each summand is non-negative. Namely, it holds:

$$U_D(1) = \sum_i \left(y_i \log\left(\frac{y_i}{x_i}\right) - (y_i - x_i) \right) \tag{7.24}$$

The sum remains unchanged because $\sum_i (y_i - x_i) = 0$. Correspondingly, 7.21, 7.8, and 7.9 hold for the equations:

$$U_i = \sum_j \left(y_{ij} \log\left(\frac{y_{ij}}{x_{ij}}\right) - (y_{ij} - x_{ij}) \right) = \sum_j x_{ij}(\rho_{ij} \log(\rho_{ij}) - (\rho_{ij} - 1)) \tag{7.25}$$

$$U_D(1) = \sum_i x_i \varrho_i \log\,(\varrho_i) = \sum_i x_i(\varrho_i \log\,(\varrho_i) - (\varrho_i - 1)) \tag{7.26}$$

$$U_D(1) = \int \varrho \log\,(\varrho)\, d^3r = \int (\varrho \log\,(\varrho) - (\varrho - 1))\, d^3r \tag{7.27}$$

The non-negativity of the summands $x_i(\rho_i \log\,(\rho_i) - (\rho_i - 1))$ follows from the non-negativity of the function $x \mapsto x\log(x) - (x-1)$, because this has value 0 at $x = 1$ and, because it is a primitive of the logarithm $x \mapsto \log(x)$, it falls monotonically for $x < 0$ and rises monotonically for $x > 0$.

The modification of the entropy definition can be seen by contrasting the graphs of the two entropy functions $x \mapsto x\log(x)$ and $x \mapsto x\log(x) - (x-1)$:

---

[4] According to inequality 7.20

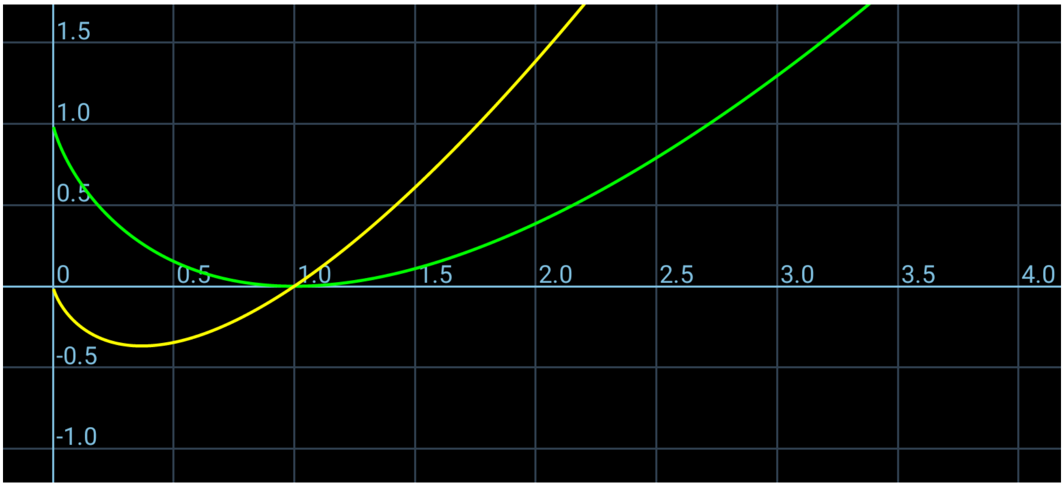


Figure 7.1: The entropy functions $x \mapsto x\log(x)$ (yellow) und $x \mapsto x\log(x) - (x-1)$ (green)

For a path $A_n := C_0C_1C_2\cdots C_n$ of a partitioning hierarchy, let $x(A_n)$ be the volume of cell C_n and $\mathcal{S}_n$ be the set of all paths of the n-th scale, which by section 7.5.1 bijectively maps the set of cells of this scale. With the densities ρ_i of the cells C_i relative respectively to the cell C_{i-1}, let be further the partially negative local entropy

$$U_-(A_n) := \rho_1\rho_2\cdots\rho_n\log(\rho_1\rho_2\cdots\rho_n) \tag{7.28}$$

and the non-negative local entropy

$$U_+(A_n) := \rho_1\rho_2\cdots\rho_n\log(\rho_1\rho_2\cdots\rho_n) - (\rho_1\rho_2\cdots\rho_n - 1) \tag{7.29}$$

of the path A_n and the cell C_n, respectively. Thus

$$\sum_{A_n\in\mathcal{S}_n} x(A_n) = 1 \tag{7.30}$$

$$U_D(1,n) = \sum_{A_n\in\mathcal{S}_n} x(A_n)U_-(A_n) = \sum_{A_n\in\mathcal{S}_n} x(A_n)U_+(A_n) \tag{7.31}$$

such that the global entropy $U_D(1,n)$ is the average of the local entropies $U_-(A_n)$ or $U_+(A_n)$ and in the limit $n\to\infty$ the following relations hold:

$$1 = \lim_{n\to\infty}\sum_{A_n\in\mathcal{S}_n} x(A_n) \tag{7.32}$$

$$U_D(1) = \lim_{n\to\infty}\sum_{A_n\in\mathcal{S}_n} x(A_n)U_-(A_n) \tag{7.33}$$

$$U_D(1) = \lim_{n\to\infty}\sum_{A_n\in\mathcal{S}_n} x(A_n)U_+(A_n) \tag{7.34}$$

The meaning of the local U_+-entropy is that it shares the typical feature of the global U-entropy, namely non-negativity, with the latter, so that through it a non-negative entropy can be assigned to each cell or point in space, whose sum or integral coincides with the global entropy.

7.5 Rational and irrational paths

7.5.1 General definition of paths

Let an infinite partitioning hierarchy in space be given, such as that implied by table 2.6 in section 2.2.2, in which each formed subcell is itself decomposed into at least two and at most finitely many subcells. We can also call the hierarchy a cell hierarchy, in which the overall system is the root of the tree formed by the hierarchy. We can further, starting at the root cell, recursively select exactly one subcell at a time to define a finite or infinite path through the hierarchy. Thus, a finite path A_n of length n is a sequence $A_n := C_0C_1C_2 \cdots C_n$ of cells such that C_0 is the entire system, C_1 is a subcell of the entire system defined by the hierarchy, C_2 is a subcell of C_1, and so on. Correspondingly, an infinite path A is an infinite sequence $A := C_0C_1C_2 \cdots$ of cells recursively selected in this way.

A finite path A_n is also an address of its last cell C_n, so we can identify the finite paths and the cells of the hierarchy with each other. For example, in table 2.6, path $C_1C_{12}C_{121}$ is the address of cell C_{121}. Moreover, for path A_n, index n is identical to scale n, the n-th level of the hierarchy.

Finally, by analogy, we can define local paths that do not start with the entire system, but with any cell C of the hierarchy. Local paths are thus defined relative to a cell C.

7.5.2 Recursive procedure of choice

Definition

A recursive procedure of choice – or function of choice – for an infinite partitioning hierarchy is a mapping that assigns to each cell C of the hierarchy one of its subcells – as a unique successor.

In the context of density entropy, a procedure of choice can be defined in the following way: Let C be any cell of the hierarchy and

$$\mathcal{A}(C, n)$$

be the set of all local paths $A_n := CC_1C_2 \cdots C_n$ of length n relative to C. Let further $\rho_1, \rho_2, \cdots \rho_n$ be the relative densities of cells $C_1, C_2 \cdots$ along such a path A_n and

$$\rho(A_n) := \rho_1\rho_2 \cdots \rho_n$$

be the density of the path relative to C.

Further, we fix an $\varepsilon_1 > 0$ uniform for the procedure of choice and first choose among all finite local paths relative to C a path A with shortest length n such that

$$\forall B \in \mathcal{A}(C, n) : \rho(B) + \varepsilon_1 < \rho(A)$$

is valid.

Thus, A is the shortest path relative to the cell C, which is sufficiently distinct from all other paths of the same length due to its greater density.

To conclude the definition of the procedure of choice for the successor of cell C: With the unique selection of a path $CC_1C_2 \cdots C_n$, C_1 can now be assigned – also uniquely – to cell C.

The procedure is not defined for a cell exactly if the described selection of a shortest path with greatest density is not possible, which in turn is exactly the case if the density in the cell is uniformly distributed.

Alternative definition

In an alternative definition for a recursive selection procedure based on spacial density fluctuations, we require that for each cell already one of its (immediate) subcells has a sufficiently larger density than the remaining (immediate) subcells. The shortest path required above then always has length 1.

The definition is a tightening, which at first is not realistic, but then is equivalent to the first one, if one does not allow any arbitrary partitioning hierarchy, but only those whose cells are resolved by their (immediate) subcells to such an extent that one of them is sufficiently distinguished by its density. Conversely, for a given hierarchy, this means that it is redefined for a cell C, if necessary, and its decomposition into subcells is refined until one of the newly emerging subcells can finally be selected with sufficient distinctness on the basis of its high density. The adjustment of the hierarchy is possible after the first definition of the selection procedure exactly when this in turn defines a selection for C.

Thus, according to this alternative definition, for each cell C there exists a subcell C_m with a largest density ρ_m relative to C, such that for all remaining subcells C_i with $i \neq m$ and their densities ρ_i relative to C:

$$\rho_i + \varepsilon_1 < \rho_m \tag{7.35}$$

Furthermore, we assume that the relative volume x_m of the cells C_m does not become too large. That is, there is a $\varepsilon_2 > 0$ such that for all cells with largest relative density:

$$x_m < 1 - \varepsilon_2 \tag{7.36}$$

The two conditions for ρ_m and x_m have – including the lower bounds ε_1 and ε_2 – no meaning for finite scales, for which namely the weaker requirements $\rho_i < \rho_m$ and $x_m < 1$ are sufficient. But they have their meaning if the limit to infinite scales is taken. Here they correspond to a reversal of the continuity principle[5], according to which what is true of the finite is also true of the infinite. The inversion consists in turning the principle into a requirement, so that what holds in the finite is required to also hold in the infinite, namely, the uniqueness when choosing a subcell as a result of its unambiguous distinguishability from the other subcells by its greatest density. For otherwise the relative densities along an infinite path $C_0C_1C_2\cdots C_n, C_{n+1}\cdots$ can evolve in such a way that, although for each cell C_n the choice of the successor cell C_{n+1} is unique, this does not hold for the limit $n \to \infty$.

Consequences for the largest relative density

Let a partitioning hierarchy be given with a procedure of choice according to the alternative definition above. For a cell C of the hierarchy, let ρ_m be the largest density relative to C, as in the last section. Using ε_1 and ε_2, also introduced in the last section, and

$$\epsilon := \varepsilon_1\varepsilon_2$$

then applies to all cells in the hierarchy:

$$\rho_m > 1 + \epsilon \tag{7.37}$$

To show this, let any cell C be given and let $\rho_1, \rho_2, \rho_3 \cdots$ be the densities of the subcells relative to C, and of these again let ρ_m be the largest. By definition of the procedure, we have:

$$\forall i \neq m : \rho_i < \rho_m - \varepsilon_1$$

---

[5] https://en.wikipedia.org/wiki/Law_of_continuity

Because $\sum_{i\neq m} x_i = 1 - x_m$, it then follows for the density mean:

$$1 = \sum_i x_i \frac{y_i}{x_i} = \sum_i x_i\rho_i = \sum_{i\neq m} x_i\rho_i + x_m\rho_m < (1-x_m)(\rho_m - \varepsilon_1) + x_m\rho_m = \rho_m - \varepsilon_1(1-x_m)$$

from which, together with the volume constraint $x_m < 1 - \varepsilon_2$ established in the last section, we finally obtain the assertion:

$$\rho_m > 1 + \varepsilon_1(1-x_m) > 1 + \varepsilon_1\varepsilon_2$$

7.5.3 Rational paths

Definition

Let be given an infinite partitioning hierarchy together with a procedure of choice. Rational paths are defined by assigning to each finite path $A_n := C_0C_1C_2\cdots C_n$ an infinite – rational path A – as follows:

- A_n is part of the rational path A, so $A = C_0C_1C_2\cdots C_nC_{n+1}C_{n+2}\cdots$.
- For all $i > n$, C_i is the cell assigned to cell C_{i-1} by the procedure of choice.

Of the two – equivalent – procedures of choice defined above, we choose the alternative definition because of its simplicity. Thus, the uniquely determined subcell with the largest relative density is selected in each case. For each rational path $A = C_0C_1C_2\cdots$ there then exists a smallest index $n \geq 0$ such that either $n = 0$ or C_n does not have the largest density, but all cells C_i for $i > n$ do. I call the local path $C_nC_{n+1}C_{n+2}\cdots$ the **period** of the rational path A, because the set $\{C_n, C_{n+1}, C_{n+2}, \cdots\}$ forms an equivalence class within the set of all cells of the hierarchy, such that a common characteristic is given to the elements of the class which distinguishes them from all other cells. Namely, the periods define equivalence classes because two periods are either identical or their sets are disjoint, and each cell of the hierarchy lies on the period of a rational path, because each cell also marks the end of a finite path, and, starting from this path, it defines a rational path whose period contains the said cell. Altogether, then, the periods of the rational paths define in their entirety a decomposition of the set of all cells of the hierarchy into equivalence classes.

Growth of absolute density

The relative densities $\rho_{n+1}, \rho_{n+2}, \cdots$ of a rational path along its period $C_nC_{n+1}C_{n+2}\cdots$ are bounded below by a value $1 + \epsilon > 1$ according to the inequality 7.37. Thus, for $m \geq n+1$:

$$\rho_m \geq 1 + \epsilon$$

Therefore, the absolute densities $\varrho_{n+1}, \varrho_{n+2}, \cdots$ grow exponentially with the beginning of the period from the $n+1$-th scale, so that for $m \geq n+1$ holds:

$$\varrho_{n+m} \geq \rho_1\rho_2\cdots\rho_n(1+\epsilon)^m \tag{7.38}$$

From this follows in particular the infinity of the absolute density ϱ of a rational path:

$$\varrho = \lim_{n\to\infty} \varrho_n = \infty \tag{7.39}$$

Moreover, the density is also infinite if ϵ is an infinitesimal in the sense defined by Eqn. 3.26 f., because it holds

$$(1+\epsilon)^m > 1 + \epsilon \cdot m$$

so that with $\epsilon := \epsilon_0, \epsilon_1, \epsilon_2 \cdots$ it follows according to the definition of the infinitesimal:

$$\varrho \geq \rho_1 \rho_2 \cdots \rho_n \lim_{m\to\infty} \epsilon \cdot m = \rho_1 \rho_2 \cdots \rho_n \lim_{m\to\infty} \epsilon_m \cdot m = \infty$$

7.5.4 Irrational paths

An infinite path is called irrational if it is not rational.

However, it is not directly apparent that they exist at all. But their existence follows from that of the rational paths, as will become clear shortly. Irrational paths are exclusively other-referential. That means, they are not only defined – as just happened above – as a complementary set with reference to the set of rational paths, but the definition of each concrete irrational path takes place exclusively by reference to rational paths, while rational paths with the beginning of their period reference only themselves, because they carry the characteristic, which uniquely defines them and by which they can also be identified, continuously before themselves. The characteristic of an irrational path, on the other hand, is that it has no infinite period of cells, each with maximum density. Instead, for no cell of the hierarchy – in its potential as a cell on irrational paths – any relative density, from the smallest to the largest, is excluded, which, in contrast to rational paths, is an expression of a statistical symmetry on the set of irrational paths.

To describe this in another way: Since the periods of the rational paths include all cells of the hierarchy, every cell of an irrational path A also lies on the period of a rational path B, and A, starting from this cell, can follow B arbitrarily far down the scale, but must eventually separate from B on a finite scale, because otherwise A would also be rational and even identical with B. It follows that an irrational path accompanies infinitely many rational paths on finite sections each, and thus references them. The length of each section is ≥ 1, it is equal to 1 in case of immediate change, i.e. if A leaves the path B already at the beginning of the period of B. The probability for immediate change is equal to 1/2 in the case of a binary partitioning hierarchy. The references to infinitely many rational paths, in turn, also imply an infinitely frequent *change* of reference, and each change is associated with the selection of a subcell whose relative density is not maximal, but for which there are no constraints besides. In fact, there is even *no* restriction for the selection of the subsequent cell from the subcells of a cell along an irrational path, because the path can also follow any rational path even arbitrarily far, so that there is also no restriction for the relative density following in each case.

7.5.5 Cardinality of the sets of irrational and rational paths

From what has been said, the size relation between the two complementary sets, the rational paths on the one hand and the irrational paths on the other hand, also follows. Namely, for any cell C of the hierarchy with $m(C)$ subcells, it holds that it is either the beginning or the continuation of the period of exactly one rational path, and the continuation of this period occurs by selecting exactly one subcell of C as the succeeding cell in accordance with the procedure of choice, while irrational paths through C can lead through any of the $m(C)$ subcells. One can use this relation to count the paths, starting at the root cell, down the scale, or to bring the growth of these numbers into the said size relation of the irrational to the rational paths, because there is a twofold exponential growth[6]:

Rational paths: The number of rational paths grows exponentially because – initially – while each cell C is the beginning or continuation of only one rational path by selecting exactly one subcell, then all the remaining of the $m(C)$ subcells are in turn the beginning of a rational path.

---

[6] The twofold exponential relation $e^{-n^2} = (e^{-n})^n$ also characterizes the normal distribution

Irrational paths: The number of irrational paths grows exponentially in relation to the number of rational paths, because for the continuation of each rational path by selecting exactly one from the $m(C)$ subcells instead $m(C)$ irrational paths emerge from the cell C.

Finally, it follows for the magnitude relation: the cardinality[7] of the irrational paths relates to the cardinality of the rational paths as the latter does to the number 1.

This corresponds to what connects the cardinalities of the irrational and rational *numbers*. The rational numbers are first defined in mathematics by the fractions m/n, for which m and n are natural numbers. That the fractions can be mapped bijectively onto the rational paths defined here, and thus subjected to a common count together with the irrational paths, will be proved in the second volume on the continuum. If, as usual, the cardinality of the rational numbers, which coincides with that of the natural numbers $\mathbb{N}$,[8] is denoted by $\aleph_0$ and that of the continuum, i.e. the irrational and real numbers, by $\mathfrak{c} = \aleph_1$, then the above relations correspond to the relation:

$$1 < \aleph_0 < \mathfrak{c}$$

7.5.6 Double reference of space and matter

The definition of rational paths given in section 7.5.3 is initially a mathematical one, but it is then extended to a physical definition based on spacial density fluctuations of matter. The question must be answered what is the meaning of the definition in the context of physics:

Usually, even in physics, we assume that space simply is – and matter is distributed in it, so that matter references space and not vice versa space references matter. An exception is the theory of relativity, which considers space, i.e. its curvature, as a function of the distribution of matter. However, this dualism unfolds only on large scales and with large masses. It is with it also a phenomenon of the extensive space. Likewise the finiteness of the speed of light belongs to the domain of the extensive space. Independently however of the scale, there is no location without reference to the distribution of matter, on large scales for example that of the fixed stars in the universe. On the scale in which we move, it is mountains, clouds, trees, houses, walls of houses, door frames, table tops, edges, markings on length scales or the paper, and so on. Now there is no reason to assume that this dependence should not exist on small scales. At most, one could object that the need to refer to matter associated with addressing in space is an anthropogenic problem alone. One would assume with it that also the empty space has the possibility to address a point of space or also a large or small environment. But this is contradicted by the symmetry of the empty space. I.e. it requires a symmetry breaking by selection of at least one point, in relation to which a location specification is not only possible, but is defined at all. This symmetry breaking is according to experience the consequence of the non-uniformly distributed matter in the position space. It is not the consequence of electromagnetic waves, which also break the symmetry of space, but which do not set thereby a reference point.

Altogether, it is therefore reasonable to assume that the hatching given to space by the distribution of matter is the basis of the definition of position and thus the basis of addressing in space also for the laws of physics, so that space and matter fundamentally define each other in a double-referential dependence. Now the addressing in space can be extensive or intensive. The extensive addressing e.g. in connection with a Cartesian coordinate system has always an intensive part, as soon as the demanded accuracy of the addressing is larger than the fixed unit of length allows. The intensive addressing – for an extensively finite system – starts at the total system and continues by recursive division. As long as the continuation of the division is finite, the addressing remains inexact. Therefore, if the dynamics of physics in the context of Newtonian

---

[7] https://en.wikipedia.org/wiki/Cardinality

[8] https://proofwiki.org/wiki/Rational_Numbers_are_Countably_Infinite

mechanics or quantum mechanics requires exact addressing for e.g. the Schrödinger equation, one must assume that the division is completed by taking the limit to infinite resolution, which only transforms the arbitrariness and indeterminacy of the finite division into definiteness.

Addressing in or on the basis of a partitioning hierarchy is only possible if the paths in the hierarchy are distinguishable, which finally are the addresses. For this again, as the preceding sections have shown, the definition of a procedure of choice for the selection of exactly one subcell to each cell is sufficient, because from it follows the definition and existence of the rational and then of the irrational paths, whose union forms the set of all paths and thus the set of all addresses, incidentally also that of the finite paths and cells, because each rational path starts in exactly one cell and vice versa exactly one rational path starts in each cell. Altogether, therefore, especially the existence and meaning of the rational paths is justified by the demand of certain addressing in space. Finally, the fact that the distribution of matter in space is the basis for the procedure of choice reflects one part of the double reference between the two mentioned above. In this context, the requirements 7.35 and 7.36 guarantee the distinctness of the rational paths and thus that of all infinite paths. Therefore, the precise addressing in space is possible if there are sufficiently strong density fluctuations even at infinite scales.

7.6 Convergence and boundedness

7.6.1 Meaning of convergence and boundedness

The inequality 7.20 shows that the U-entropy $U_D(1, n)$ grows monotonically as a function of the scale n of a partitioning hierarchy, and therefore the convergence of the U-entropy is insolubly related to its growth. That is, it converges exactly when it is bounded[9].

In this context, the study of density entropy in section 3.5.3 with the result – for a partitioning hierarchy equal to the dual system – in inequality 3.73 shows that density entropy has the potential to fall logarithmically infinitely, just as Shannon entropy grows logarithmically infinitely. Since the U-entropy arises from the normalized density entropy by sign reversal, the U-entropy can also grow infinitely and is similar to the Shannon entropy in this respect.

Therefore, everything indicates that bounded growth of entropy is crucial for its theoretical meaning, because without boundedness and thus without convergence, entropy is not only indeterminate, but undefined, which calls the concept of entropy as a whole into question.

7.6.2 Boundedness due to finiteness of hierarchy

The simplest way to bring about boundedness is to assume a smallest scale below which there is only uniformly distributed density, which corresponds to the *coarse graining* mentioned in section 3.5.2 in the context of Gibbs entropy, but contradicts the precise position determination by exact addresses as demanded in section 7.5.6.

For the U-entropy given by Eq. 7.10

$$y_i \log\left(\frac{y_i}{x_i}\right) + y_i \sum_j y_{ij} \log\left(\frac{y_{ij}}{x_{ij}}\right)$$

this means that if the index i marks the cells C_i of a flat partitioning at the smallest scale, the entropies

$$\sum_j y_{ij} \log\left(\frac{y_{ij}}{x_{ij}}\right)$$

---

[9] https://en.wikipedia.org/wiki/Monotone_convergence_theorem

vanish for all cells C_i regardless of their decomposition into subcells C_{ij}.

In consequence of such finiteness, space is discrete with all further consequences of indeterminacy arising from it.

An argumentation according to which a smallest scale is set in phase space by Heisenberg's uncertainty principle means that the density entropy and the motion entropy are not only related in their temporal evolution, but are also only defined together. So there would not be then the definition of a density entropy separate from that of the entropy of motion, but only one. Also the uncertainty relation is valid for a single particles and not for the cumulative occupation probabilities y_i of a particle collective.

7.6.3 Boundedness due to bounded absolute density

Another simple way to limit entropy growth is to strictly limit the absolute density for all paths. For this the mean property of entropy is the foundation. Namely, if ϱ_{max} is an upper bound on the absolute density, then according to Eqn. 7.8 for the flat partition corresponding to a partitioning hierarchy holds:

$$U_D(1) = \sum_i y_i \log\left(\frac{y_i}{x_i}\right) \leq \sum_i y_i \log(\varrho_{max}) = \log(\varrho_{max}) \tag{7.40}$$

However, this strict limitation on the absolute density is incompatible with the infinite density of rational paths according to equation 7.39 and thus contradicts the existence of uniquely distinguishable infinite paths, insofar as they are defined on the basis of density fluctuations and a procedure of choice according to section 7.5.2.

On the other hand, the condition seems – at first sight – to allow quite well fluctuations of the relative densities of finite size even in the limit to infinite scales. For an infinite path with relative densities $\rho_1, \rho_2, \rho_3 \cdots$, the sequence of absolute densities with values $\varrho_n = \rho_1\rho_2\rho_3 \cdots \rho_n$ is – in accordance with the condition 7.40 – not necessarily required to converge, but only be bounded and satisfy the condition $\varrho_n \leq \varrho_{max}$, for which convergence is not mandatory. The simple extreme case, however, where for each cell among its subcells there is one whose relative density does not fall below a certain globally valid value $1+\epsilon$, shows that the condition does set limits to such density fluctuations existing on all scales. Because in this case for each cell a path can be defined, which leads through the cell and whose subsequent cells are in each case those, whose relative density satisfies $\geq 1+\epsilon$. These paths are finally again rational paths with infinite absolute density. In other words, the relative densities as a function of scale must converge to 1 – at least in the bulk – for the absolute densities to be undoubtedly bounded.

In particular, incompatible with the density-based definition of infinite paths are continuous density functions, since by definition they do not distinguish points of space sufficiently clearly from other points of their neighborhood.

Separately, it is interesting to consider the consequences of an analogous constraint on Shannon entropy. Such a constraint is due to $-\sum p_i \log(p_i) = \sum p_i \log(1/p_i)$, with an upper bound $c > 0$, identical to the condition $1/p_i \leq c$, from which follows:

$$p_i \geq p_{abs,min} := \frac{1}{c}$$

If one associates the probability p_i, like the y_i of the U-entropy, with the particle occupation probability in cell C_i of size x_i, then the condition means that above a certain scale the further division of the cell C_i is not accompanied by a corresponding division of the probability p_i. Thus, for a division of C_i into subcells C_{ij} with relative sizes x_{ij} and particle shares p_{ij}, for exactly one subcell C_{ik} the particle share satisfies

$$p_{ik} = p_i$$

and for all others $j \neq k$:

$$p_{ij} = 0$$

If the division is infinitely continued, the condition defines exactly one infinite path A leading through cell C_i that has particle share p_i, while all other infinite paths through C_i have particle shares 0. A is then the address of a mass point of classical mechanics.

7.6.4 Boundedness by infinitesimals

Boundedness with bipartition

We now assume the existence of a procedure of choice according to section 7.5.2, which is the basis for the density-based definition and distinguishability of infinite paths. Under this assumption, according to inequality 7.37, the largest relative density is bounded below by a value $1 + \epsilon$, from which in turn follows, according to section 7.5.3, the exponential growth of the absolute densities of rational paths and hence the infinity of these densities, even when ϵ has the meaning of an infinitesimal. The premises of the last section are thus certainly violated. On the other hand, the rational paths have such a small share among all paths that they need not necessarily stand in the way of the finiteness of entropy.

However, it will be shown below that the U-entropy for a non-infinitesimal ϵ grows linearly as a function of scale in the same way as the Shannon entropy and thus infinitely, which corresponds to a logarithmic growth as a function of the partition size, i.e., the number of cells on the scales. In this context it can be mentioned, because thus both entropies have in principle this growth equally, that the logarithmic growth is of extraordinary slowness, as it is illustrated by the following graph of the logarithm function:

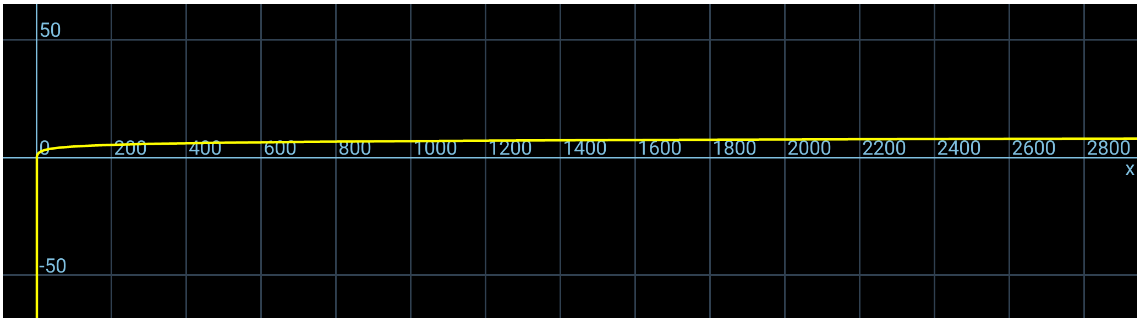


Figure 7.2: Der Graph des Logarithmus $x \mapsto \log(x)$ zur Illustration seines geringen Wachstums

It is true that the graph is deceptive in that it is plotted against the background of a deliberately small scale. On the other hand, however, it is exactly what determines an essential part of the discussion about entropy, namely taking the limit to even infinitely small scales. As will be proved in the second volume on the continuum, the growth of the logarithm is even so slow that the sequence

$$\log(1), \log(2), \log(3), \cdots \log(n), \cdots$$

on the one hand grows infinitely though, since there is no upper bound or at least it cannot be named, but on the other hand the sequence does not traverse the set of natural numbers $\mathbb{N}$, so that

$$\lim_{n\to\infty} \log(n) < \lim_{n\to\infty} n$$

is true. Therefore, the conjecture can be made that logarithmic growth has the fundamental potential for finiteness as well as infinity. For Shannon entropy, the section 3.3.3 has revealed the side of infinite growth even when the non-uniform distributions that keep Shannon entropy small are perfect except for an infinitesimal. The reverse is true for U-entropy, since it becomes small as the uniform distribution increases. The ambiguity of logarithmic growth is now seen in the fact that U-entropy is bounded when the density is almost perfectly, i.e., uniformly distributed except for an infinitesimal, which allows the infinite paths – through the infinitesimal – to be distinguishable.

For the upper bound estimation now of the entropy growth, we assume in this section, analogously to section 3.3.3 on the infinite growth of Shannon entropy and as already occurred in section 3.5.3 to investigate the intensivity of density entropy, a simple binary partitioning hierarchy in which, at each scale, each cell C_i is decomposed into exactly two subcells with the same relative sizes $x_{i1} = x_{i2} = 1/2$. Let the particle shares of the two cells be $y_{i1} = 1/2 - \xi$ and $y_{i2} = 1/2 + \xi$, respectively, so that ξ is a measure of the size of the non-uniform distribution of particles – here over the two halves – as in Shannon entropy. For $\xi = 0$ they are uniformly distributed between the halves, for $\xi = 1/2$ and $\xi = -1/2$ they are entirely either in one or the other half.

For the entropy U_i associated with the bipartition of the cell C_i, which we now call ΔU_i because of the accentuation of the increase, according to Eqn. 7.21

$$\Delta U_i = x_{i1}\rho_{i1}\log(\rho_{i1}) + x_{i2}\rho_{i2}\log(\rho_{i2}) \tag{7.41}$$

holds and for the relative densities ρ_{i1} and ρ_{i2}:

$$\rho_{i1} = \frac{1/2-\xi}{1/2} = 1 - 2\xi, \rho_{i2} = \frac{1/2+\xi}{1/2} = 1 + 2\xi \tag{7.42}$$

So also applies:

$$\Delta U_i = \frac{1}{2}[(1-2\xi)\log(1-2\xi) + (1+2\xi)\log(1+2\xi)] \tag{7.43}$$

It is worthwhile to compare this expression to that in Eqn. 3.21 for the increase in Shannon entropy

$$\Delta S_i = \log(2) - \frac{1}{2}((1-2\xi)\log(1-2\xi) + (1+2\xi)\log(1+2\xi))$$

because it becomes clear that the U-entropy literally turns the characteristics of the Shannon entropy upside down, as is also shown in the following figure with the graphs for both functions, ΔU_i and ΔS_i:

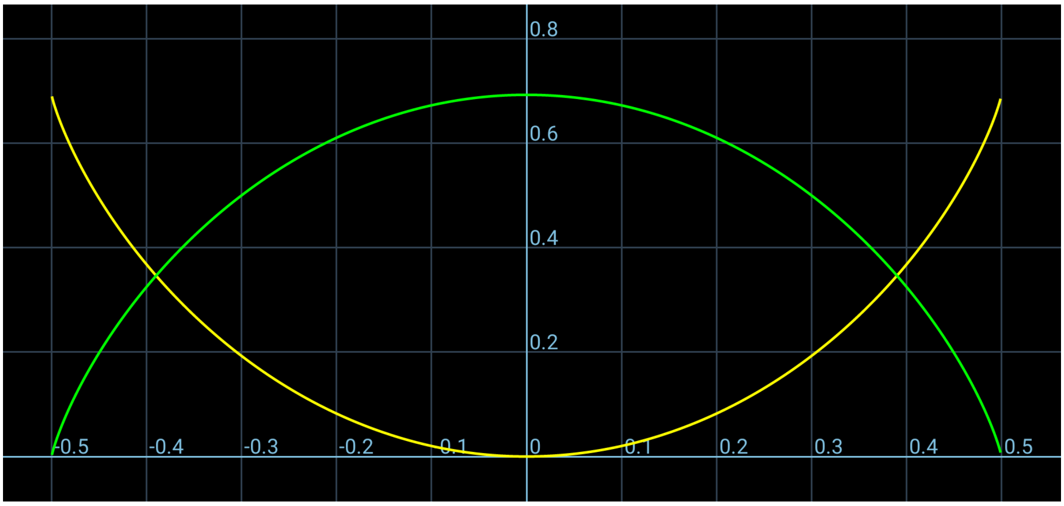


Figure 7.3: ΔU_i (yellow) and ΔS_i (green) for the bipartition as a function of the non-uniform distribution measure ξ.

Both entropies require values close to 0 for their global boundedness, i.e., it must be $\Delta U_i \approx 0$ or $\Delta S_i \approx 0$. The most important difference between them is that for ΔU_i this is the case for uniform distribution in the neighborhood of $\xi = 0$ with derivative 0 and for ΔS_i it is the case for non-uniform distribution in the neighborhoods of $\xi = 1/2$ and $\xi = -1/2$ with derivatives $\pm\infty$.

To perform the estimation, we note that of the densities of the two halves ρ_{i2} is the larger and hence the largest, so that, because we assume the existence of a procedure of choice, it is subject to the condition 7.37. That is,

$$1 + 2\xi = \rho_{i2} \geq 1 + \epsilon$$

wherefrom

$$\xi \geq \frac{\epsilon}{2}$$

follows. Putting this into Eqn. 7.43 above for ξ, then it follows further, analogously to Eqn. 3.23 ($\Delta S_i \geq \Delta S_{min} := -2\epsilon \log(\epsilon)$), from the strict monotonicity of the entropy increase ΔU_i for $\xi > 0$

$$\Delta U_i \geq \Delta U_{min} := \frac{1}{2}[(1-\epsilon)\log(1-\epsilon) + (1+\epsilon)\log(1+\epsilon)] \tag{7.44}$$

and from this according to Eqn. 7.19 for the entropy increase $\Delta U_D(1,n)$ from one scale $n-1$ to the next n:

$$\Delta U_D(1,n) = \sum_i y_i \Delta U_i \geq \Delta U_{min} \tag{7.45}$$

For a finite lower bound $\epsilon > 0$, together with Eqn. 7.20, this results in the same way as for the Shannon entropy[10] in the infinity of the U-entropy:

$$U_D(1) = \sum_{n=1}^{\infty} \Delta U_D(1,n) \geq \lim_{n\to\infty} \Delta U_{min} \cdot n = \Delta U_{min} \lim_{n\to\infty} n = \infty \tag{7.46}$$

However, the difference in entropies as a function of the non-uniform distribution measure ξ according to figure 7.3 gives a clear indication that the estimate for ΔU_i by the minimum entropy ΔU_{min} based on an infinitesimal ϵ may be different. The direct graphical comparison of the minimum entropies ΔU_{min} and $\Delta S_{min} = -2\epsilon \log(\epsilon)$ is even more evident:

---

[10] Cf. the relation 3.25

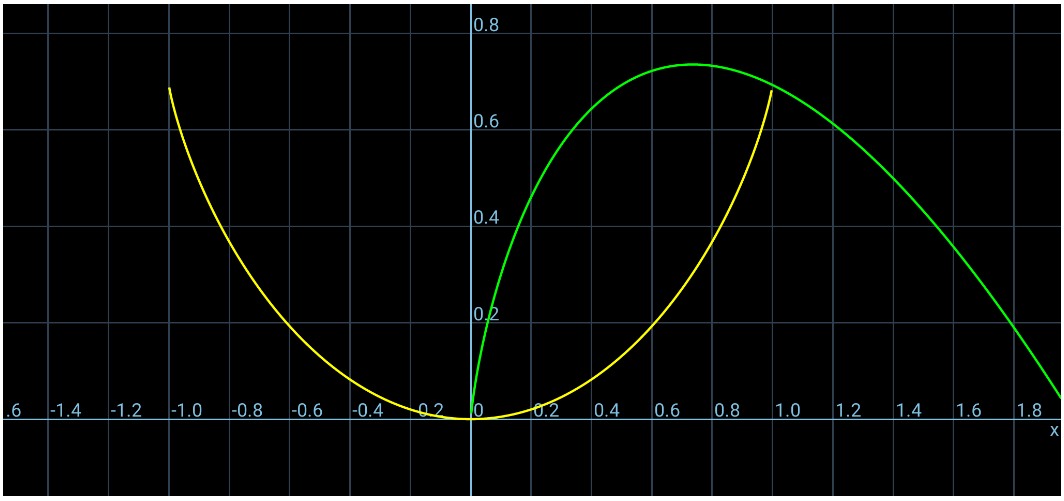


Figure 7.4: Minimal ΔU_{min} (yellow) and ΔS_{min} (green) for the bipartition

The comparison of the minimum entropies is clearer because the critical point with entropy 0, namely $\epsilon = 0$, is the same for both entropies. Otherwise, however, their behavior at this point is virtually opposite. Besides the differences already noted for the figure 7.3 – infinite derivative for ΔS_{min} at $\epsilon = 0$ in contrast to vanishing derivative for ΔU_{min} – it becomes clear that ΔU_{min} is also defined for $\epsilon < 0$.

For the estimation now based on an infinitesimal[11] one can evolve ΔU_{min} above in second order around $\epsilon = 0$ and obtain:

$$\Delta U_{min} = \frac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3)$$

Replacing ϵ by an infinitesimal $\epsilon = \epsilon_1, \epsilon_2, \epsilon_3 \cdots$ one finally obtains instead of 7.46, if one chooses for the infinitesimal ϵ the sequence with members $\epsilon_n = C/\sqrt{n}$ for which C is a constant:

$$U_D(1) = \sum_{n=1}^{\infty} \Delta U_D(1,n) \geq \lim_{n\to\infty} \Delta U_{min} \cdot n = \lim_{n\to\infty} n \cdot \left(\frac{1}{2}\epsilon_n^2 + \mathcal{O}(\epsilon_n^3)\right) = \frac{1}{2}C \tag{7.47}$$

Thus it is shown that the U-entropy does not necessarily grow infinitely, if the relative density also varies on the infinite scales with at least infinitesimal size. A necessary condition for the boundedness and thus the convergence of the entropy is thus fulfilled. Now, in the same way, one can define plausible conditions whose fulfillment also implies the existence of an upper bound for the entropy. For example, the assumption that the fluctuations of the relative densities are not only globally bounded downward by an infinitesimal, but in any case below an arbitrarily small scale also bounded upward.

Boundedness in the general case

To consider the general case, instead of the expression for the entropy increase ΔU_i in Eqn. 7.41 we choose the expression given by Eqn. 7.25

$$\Delta U_i = \sum_j x_{ij}(\rho_{ij}\log(\rho_{ij}) - (\rho_{ij} - 1))$$

which is determined by the non-negative entropy function

$$f : x \mapsto x\log(x) - (x - 1)$$

---

[11] Cf. the definition in 3.27

and which is shown in Figure 7.1. Assuming, as suggested in the conclusion of the previous section, that the relative density fluctuations on small scales are small and thus the densities ρ_{ij} are even in a tight ϵ environment of 1, so that

$$|\rho_{ij} - 1| < \epsilon$$

applies, then one can evolve the function f around 1 again and obtain

$$f(1+x) = \frac{1}{2}x^2 + \mathcal{O}(x^3)$$

and thus, because of the monotonicity of f on both sides of 1:

$$\Delta U_i < \frac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3)$$

From here one can follow the reasoning of the last section on the basis of an infinitesimal and finally obtain the required estimate of the entropies $U_D(1,n)$ and thus the finiteness of the entropy $U_D(1)$.

7.6.5 Density and entropy for rational and irrational paths

Individual and collective entropy

According to Eqn. 7.31, the collective entropy $U_D(1,n)$ of all finite paths of scale n equals:

$$U_D(1,n) = \sum_{A_n \in \mathcal{S}_n} x(A_n) U_-(A_n) = \sum_{A_n \in S_n} x(A_n) U_+(A_n)$$

It is then an average of the individual entropies $U_-(A_n)$ or $U_+(A_n)$. Equivalently, it is the mean of the local entropies in the cells of the flat partitioning defined by the n-th scale of the partitioning hierarchy. Both are also true in the limit $n \to \infty$. The individual entropies of the finite paths become those of the rational and irrational paths, and the divisible cells of finite size > 0 of the finite scales become the indivisible rational and irrational points of space whose addresses are the infinite paths. Because rational and irrational paths are by definition uniquely distinguishable, the entropy of an individual rational path can also be separated from that of an irrational path, and similarly the collective entropy of all rational paths can be separated from that of irrational paths.

An individual infinite path $A = C_0, C_1, C_2, \cdots$ can be assigned an entropy alternatively according to the equations 7.28 or 7.29 so that – with the relative densities $\rho_1, \rho_2, \rho_3, \cdots$ and the absolute densities $\varrho_n = \rho_1 \rho_2 \cdots \rho_n$ along the path – for the entropies $U_-(A)$ or $U_+(A)$ holds:

$$U_-(A) = \lim_{n\to\infty} \varrho_n \log(\varrho_n) \tag{7.48}$$

$$U_+(A) = \lim_{n\to\infty} \varrho_n \log(\varrho_n) - (\varrho_n - 1) \tag{7.49}$$

Rational paths

For a single rational path, the entropy is infinite according to the two definitions 7.48 and 7.49, since the absolute density ϱ_n according to Eqn. 7.39 f. grows infinitely.

Notwithstanding this, the collective contribution of all rational paths to the total entropy is zero. To this end, we first establish that (for rational paths)

$$U_+(A) < U_-(A)$$

applies, so that an upward estimate of the collective entropy based on the individual values $U_{-}(A)$ is sufficient. To perform the estimation, we again assume a binary partitioning hierarchy with cell bipartition and further note that – taking into account the 7.5.5 section on the cardinality of rational and irrational paths – the number of rational paths counted with respect to a scale n grows $\sim 2^n$, but falls $\sim 2^{-n}$ relative to the growing number of irrational paths. The resulting small share of rational paths relative to that of irrational paths is responsible for the fact that their collective entropy share in the total entropy is also small, despite the unbounded growth of their absolute densities.

For the relative densities ρ_n, because of the assumed cell bipartition, according to the equations 7.42 holds:

$$0 \leq \rho_n \leq 2$$

If one assumes that there is a q with $1 < q < 2$ and a scale N such that for all relative densities ρ_n at all scales $n > N$ even

$$\rho_n \leq q$$

is valid, then it follows for $U_{-}(A_n)$ of a partial path $A_n = C_0, C_1, C_2, \cdots C_n$ of A with $n > N$:

$$U_{-}(A_n) = \rho_1\rho_2\rho_3\cdots\rho_n \log(\rho_1\rho_2\rho_3\cdots\rho_n) \leq 2^N q^{n-N} \log(2^n) = 2^N q^{n-N} n \log(2)$$

Let then X_n be the share of the counted rational paths in all counted paths on the scale n, for which $X_n \leq 2^{-n}$ holds according to what was said at the beginning. Then, because of $q/2 < 1$, the entropy share of the rational paths is bounded above by:

$$\lim_{n\to\infty} X_n 2^N q^{n-N} n \log(2) \leq \lim_{n\to\infty} \left(\frac{q}{2}\right)^{n-N} n \log(2) = 0$$

Irrational paths

The section 7.6.3 about the boundedness of entropy by the boundedness of absolute density shows how much the boundedness of U-entropy is connected with that of absolute density. This is true not only for the collective, but in the same way for the entropy of a single infinite path, because it is directly functionally related to the density. I.e. the entropy of the path is bounded exactly if it is the absolute densities $\varrho_1, \varrho_2, \cdots$ along itself. Downward, the restriction holds anyway – first for the absolute densities, which by definition are ≥ 0, but also for the entropies, which is directly illustrated by figure 7.1. Now for an irrational path, the bounded growth of its absolute densities cannot be guaranteed, to be sure. However, one can assume that the probability of infinite growth for an arbitrarily chosen path is zero, because for it, by definition according to section 7.5.4, the probability for each of its cells $C_1, C_2, C_3, \cdots$ has the same size for each possible relative density, e.g., for densities < 1 on the one hand and those > 1 on the other. Therefore, also its probability of infinite density growth

$$\lim_{n\to\infty} \rho_1\rho_2\rho_3\cdots\rho_n = \infty$$

and likewise its convergence to 0

$$\lim_{n\to\infty} \rho_1\rho_2\rho_3\cdots\rho_n = 0$$

equals zero. However, it does not necessarily follow that they converge. Rather, the logarithmized absolute density

$$\sum_{n=1}^{\infty} \log(\rho_n)$$

of a path represents a series which is almost always bounded, but just not necessarily convergent, even if the logarithmized relative densities $\log(\rho_n)$ form a zero sequence, because this is only a necessary, but not a sufficient condition for the convergence of the series, because in general the series, similar to an *alternatung series*[12], consists of positive and negative summands at the same time.

Now according to the conditions of section 7.6.4 the boundedness and convergence of the total entropy is certain and thus necessarily also that of the collective entropy of the irrational paths, since that of the rational paths is equal to zero. Of course, this convergence is also connected with the boundedness of the absolute densities of almost all irrational paths, which has just been justified by the probability argument, and which can also be justified in the following simple but quite different way:

Namely, this boundedness is a consequence of the density normalization

$$\sum_i x_i \varrho_i = \sum_i x_i \frac{y_i}{x_i} = 1$$

Because of this, the upward growth of the density can only be compensated by falling but bounded values because of its global lower bound 0. To see this more clearly, let the index i again mark the cells of the flat partitioning defined by a scale n of a partitioning hierarchy. Let then $X(n)$ be the cumulative volume of those cells of the scale whose absolute density ρ_i has at least a given value $\varrho(n)$ such that holds:

$$\varrho_i \geq \varrho(n)$$

Because of normalization is

$$1 \geq \sum_{i:\varrho_i \geq \varrho(n)} x_i \varrho_i \geq \sum_{i:\varrho_i \geq \varrho(n)} x_i \varrho(n) = X(n)\varrho(n)$$

and thus:

$$X(n) \leq \frac{1}{\varrho(n)}$$

So if the sequence of values $\varrho(n)$ satisfies

$$\lim_{n\to\infty} \varrho(n) = \infty$$

then the volume of the affected cells in the limit of infinite resolution is

$$\lim_{n\to\infty} X(n) = \lim_{n\to\infty} \frac{1}{\varrho(n)} = 0 \tag{7.50}$$

so that the set of paths with an infinitely increasing absolute density is a null set, insofar as one connects the size of the uncountable set of paths leading through a cell with the size of the cell, i.e., its measure.

Thus, conversely, the set of paths with bounded absolute density has measure 1. Since the rational paths do not belong to it in any case because of their infinite density and their share is also a null set anyway, the boundedness concerns therefore the irrational paths and of them almost all.

---

[12] https://en.wikipedia.org/wiki/Alternating_series

7.6.6 Entropy as its own convergence criterion

The past sections have made clear the connection between the boundedness of densities and that of entropy, and thus the connection between density and convergence of entropy. This involved necessary or sufficient conditions for the boundedness of entropy in terms of lower or upper bounds $1+\epsilon$ for the relative density, whether ϵ has the meaning of an infinitesimal or not. Now, setting bounds, in this case for densities to satisfy *without* exception, is a hard convergence criterion. Density entropy, however, does not need such a hard criterion, as shown first of all by the rational paths which do not obey the density constraint and yet do not jeopardize the convergence of entropy. According to Eqn. 7.19, the entropy increase on the n-th scale is equal to:

$$\Delta U_D(1,n) = \sum_{ij} y_i y_{ij} \log(\rho_{ij})$$

The logarithmization $\log(\rho_{ij})$ maps the ratios of densities ρ_{ij} relative to 1 to values relative to 0, and their mean is $\Delta U_D(1,n)$. At very small scales in the limit $n \to \infty$, the index i represents an uncountable set which admits any number – certainly in any case countably infinite – of exceptions to the constraint on the densities, without any influence on the mean. It follows that entropy itself – as a mean value in the form of the above equation – is the actual and considerably finer[13] convergence criterion for the boundedness, and hence the convergence of itself, than the strict boundedness of densities could ever be.

However, in this a decisive role is played additionally by the increasingly equal clustering of densities ρ_{ij} on both sides of zero as the scale decreases, which, if they are close to 1, are clustered not only in equal numbers but also with equal magnitudes on both sides of 0 because of the then valid relation $\log(\rho_{ij}) \approx \rho_{ij} - 1$, which gives the infinitesimal its meaning.

Finally, with respect to this symmetry around zero, the essential difference to Shannon entropy is shown here in a particularly striking way. For its analogue to the above equation is given by Eqn. 3.17:

$$\Delta S(n) = \sum_{ij} p_i p_{ij} \log\left(\frac{1}{p_{ij}}\right)$$

Unlike the density ρ_{ij}, however, $1/p_{ij} > 1$ is always true and thus also without exception:

$$\log\left(\frac{1}{p_{ij}}\right) > 0$$

---

[13] Cf. the end of section 3.3.4

Bibliography

[1] Sheldon Goldstein, Joel L. Lebowitz, Roderich Tumulka, , and Nino Zanghı. Gibbs and boltzmann entropy in classical and quantum mechanics. *arXiv*, 2019.

[2] Wikipedia. *Gibbssches Paradoxon — Wikipedia, die freie Enzyklopädie.* `https://de.wikipedia.org/w/index.php?title=Gibbssches_Paradoxon&oldid=213748298`, 2021. [Online; Stand 22 February 2022].

[3] Wikipedia contributors. *Boltzmann's entropy formula — Wikipedia, The Free Encyclopedia.*

[4] Wikipedia contributors. *Entropy (statistical thermodynamics) — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Entropy_(statistical_thermodynamics)&oldid=1051086306`, 2021. [Online; accessed 15 December 2021].

[5] Wikipedia contributors. *H-theorem — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=H-theorem&oldid=1047321923`, 2021. [Online; accessed 28 December 2021].

[6] Wikipedia contributors. *History of entropy — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=History_of_entropy&oldid=1051086359`, 2021. [Online; accessed 15 December 2021].

[7] Wikipedia contributors. *Parabolic partial differential equation — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Parabolic_partial_differential_equation&oldid=1061279972`, 2021. [Online; accessed 06 January 2022].

[8] Wikipedia contributors. *Quantum entanglement — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Quantum_entanglement&oldid=1062667869`, 2021. [Online; accessed 30 December 2021].

[9] Wikipedia contributors. *Well-posed problem — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Well-posed_problem&oldid=1000080927`, 2021. [Online; accessed 06 January 2022].

[10] Wikipedia contributors. *Poincaré recurrence theorem — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Poincar%C3%A9_recurrence_theorem&oldid=1068845662`, 2022. [Online; accessed 3-April-2022].