On the constitutive role of large numbers for the possibility of theory development for macroscopic systems

Reimer Kühn
Department of Mathematics, King's College London
Strand, London WC2R 2LS, UK

May 9, 2007

Abstract

The present contribution deals with the central role that large numbers have played in the process of developing theories about macroscopic systems. We begin by analysing the empirical foundations of this observation, the emergence of stability and regularity through averaging in large systems, and describe their formalisation via limit theorems of mathematical statistics. We choose to adopt a formalisation which emphasises properties of descriptions on scales which are much larger than the atomic or molecular scale at which, according to current understanding, the phenomena being described have their origin. By going on to consider the consequences that the empirical foundation of our observation has for our neural information processing apparatus, we are forced to conclude that large numbers play a crucial role already in allowing stable perception and representation of external and internal reality, and thus appear to be constitutive for all theorising about the world.

1 Introduction

Regularities observed in nature have always provided the main driving force for scientific endeavour. First and foremost, these were the regularities of day following night, the recurring seasons, or the orbits of the sun, the moon or the planets in the day and night skies.

The examples just mentioned are astronomical in nature, or caused by regularities of celestial mechanics; indeed the motions of the celestial objects that can be observed with the naked eye have been regarded as paradigms of regularity and lawfulness. It will thus not come as a surprise that astronomy has been one of the first and primary objects of natural philosophy throughout the ancient world [1].[1] And it is fair to say that astronomy once more played a crucial role in the revolutionary transition to quantitative modern science associated with then names of Copernicus, Kepler, Galileo, Newton and others.

The present essay is, however, not primarily concerned with phenomena at the large astronomical scale, nor with discoveries at the very small, the atomistic or even sub-nuclear scale. We shall instead take physical properties of macroscopic systems to provide the starting point of our discussion.

When speaking of macroscopic systems, we have things in mind that constitute our natural environment, comprising solids, liquids and gaseous systems. By *physical* properties of these systems we wish to understand properties such as their density, heat capacity, compressibility, elastic properties, colour, magnetic or electric properties, and more, as well as the way in

[1] The significance attached to regularity of celestial phenomena is still reflected in the original meaning of the Greek word nowadays used to designate the universe: *κόσμος*, meaning *order*.

which these properties depend on environmental conditions such as temperature, pressure, or electromagnetic fields.

The central hypothesis to be advanced in what follows is that our very ability of developing theories for physical properties of macroscopic systems has crucially depended on the fact that such systems are invariably composed of a vast number of atomic or molecular constituents — numbers so vast in fact that they are in a certain well-defined sense (to be explained below) indistinguishable from infinity.

The construction of proper theories for such system has been a fairly recent addition to the spectrum of scientific endeavour, beginning as it does with the advent of thermodynamics which is itself closely entwined with the early days of the industrial revolution and with the construction of steam-engines in the early 19th century.

Without belabouring the obvious it may be worth pointing out that, when mentioning physical properties of macroscopic systems we were, indeed, listing *properties*. Observable regularities of natural *processes* are not the first thing coming to mind when thinking of these properties. And the very discovery that these constitute 'explananda' — things to be explained — would seem to constitute a highly non-trivial insight in its own right. It is thus perhaps not a historical accident that discoveries of this kind began to be made with the beginning of the industrial revolution, when systematic manipulation of nature became more frequent, and with it observations of regular responses of macroscopic systems to manipulation.

In what follows we shall attempt to justify the above-mentioned hypothesis concerning the constitutive role of large numbers for the creation of theories for macroscopic systems. The key point of this justification will consist in demonstrating that very large numbers are a crucial *prerequisite* for the stability of physical properties of such systems, as well as for the reproducible way in which such systems react to changing environmental conditions. Such stability and regularity was recognised by Kant as a crucial empirical prerequisite of phenomenal consciousness:

> Würde der Zinnober bald rot, bald schwarz, bald leicht, bald schwer sein, ein Mensch bald in diese, bald in jene tierische Gestalt verändert werden, am längsten Tage bald das Land mit Früchten, bald mit Eis und Schnee bedeckt sein, so könnte meine empirische Einbildungskraft nicht einmal Gelegenheit bekommen, bei der Vorstellung der roten Farbe den schweren Zinnober in die Gedanken zu bekommen, oder würde ein gewisses Wort bald diesem, bald jenem Dinge beigelegt, oder auch dasselbe Ding bald so, bald anders benannt, ohne daß hierin eine gewisse Regel, der die Erscheinungen schon von selbst unterworfen sind, herrschte, so könnte keine empirische Synthesis der Reproduktion statt finden.[3]
>
> If cinnabar were sometimes red, sometimes black, sometimes light, sometimes heavy, if a man changed sometimes into this and sometimes into that animal form, if the country on the longest day were sometimes covered with fruit, sometimes with ice and snow, my empirical imagination would never find opportunity when representing red colour to bring to mind heavy cinnabar. Nor could there be an empirical synthesis of reproduction, if a certain name were sometimes given to this, sometimes to that object, or were one and the same thing named sometimes in one way, sometimes in another, independently of any rule to which appearances are in themselves subject[4].

As phenomenal consciousness of perceptions is itself surely prerequisite for their theoretical penetration, the hypothesis concerning the crucial role of large numbers appears to carry much further than its formulation in the initial context of the possibility of theory development for macroscopic system might suggest.[2] We shall return to this wider aspect later on when considering the role of large numbers for (neural) perception and representation.

We shall take as our starting point the premise that a probabilistic description of macroscopic systems using methods of Boltzmann-Gibbs statistical mechanics is essentially correct (in the sense of adequate). This premise has itself been a matter of some debate. Here we shall take the *pragmatic* point of view that this form of probabilistic description of macroscopic systems is empirically very well supported. To further underpin the adequacy of this description of nature let us mention the following facts: (i) on the microscopic, atomistic level a proper description of physical systems would have to be of a quantum mechanical nature and thus fundamentally non-deterministic [2];[3] (ii) even within a classical description it is fairly straightforward to see that, e.g., scattering in many body systems is chaotic in the sense that even small changes in initial conditions will generate microscopically very different configurations at later times. This observation together with the fact that irregular and, indeed, microscopically fluctuating boundary conditions such as created by continuously changeable walls of a container that may enclose a liquid or a gas are hardly ever included in their microscopic description, suggest that a probabilistic description of such systems is indeed the truly adequate one.

Starting from the premise just described, the justification for our main hypothesis will consist in (i) describing that macroscopic systems invariably consist of a vast number of constituents, and that this property is the main reason for the fact that such systems exhibit stable macroscopic properties and that they react in predictable ways to external forces and fields, (ii) to argue that such stability is according to what has been said a *key prerequisite* for acquiring knowledge of their properties and thus for attempting theoretical descriptions thereof, (iii) to describe the mathematical foundation of the observed regularities in the form of limit theorems of mathematical statistics and to relate these limit theorems with key features of large-scale descriptions of these systems, (iv) to highlight the particular role of the so-called thermodynamic limit of infinite system size in concrete formulations of theoretical descriptions, and finally (v) to apply our reasoning to the neural apparatus involved in perception and representation, thereby realising that large numbers are decisive for stable perception and representation of, and thus for theorising about the world at an even more fundamental level than our initial discussion about macroscopic systems would suggest.

2 Empirical aspect: large numbers and regularity

Let us then have a look at the empirical foundation of our hypothesis. It is related to the observation that macroscopic systems — which according to our premise are adequately characterised as stochastic systems —- would not exhibit stable macroscopic properties if they wouldn't consist of huge numbers of atomic or molecular constituents.

This observation, here formulated in colloquial terms, does have a precise meaning in the context of a description in terms of Boltzmann-Gibbs statistical mechanics. Finite systems, are according to such a description, *ergodic*. They would therefore attain all possible micro-states with probabilities given by their Boltzmann-Gibbs equilibrium distribution, and would therefore in general *also* exhibit fluctuating *macroscopic* properties

To discuss an example: a piece of diamond might disintegrate into a heap of carbon dust,

[2] I am indebted to Christel Fricke for pointing out that my arguments may be accentuated along this line, and for drawing my attention to the pertinent Kant source.

[3] Without further commitment as to whether this description would or would not be *complete*, no metaphysical point of view in the sense of stating, or rejecting a fundamental indeterminacy of nature is implied. The author admits that he doesn't know whether or not God plays dice. Moreover, he believes that, which of these alternatives were true, would be immaterial for the reasoning in the present paper.

which could spontaneously reassemble as graphite, which might in turn be transformed into a piece of pit-coal, thereafter perhaps back into a diamond, and we haven't even begun to mention the more exotic forms of carbon, such as the various fullerenes, or carbon-nanotubes and so on that we would also have to reckon with. If such transformations would only occur sufficiently frequently, we would not have reasons to even just *talk* about diamond, carbon dust, pit-coal, or the more exotic forms of carbon.

The sole reason for the fact that such a scenario does not correspond to our everyday experience is in the fact that macroscopic matter invariably consists of huge numbers of atomic or molecular constituents. Even if — staying with our example — we are used to imagine small quantities when thinking of diamond, we would normally have to consider samples containing at least 10^{20} carbon atoms.[4]

Although the numbers just mentioned are clearly *finite*, they are surely unimaginably large, and it is the fact that they are so very large which is responsible for the fact that transitions between different manifestations of macroscopic matter are sufficiently rare to ensure stability and regularity of the various different manifestations. Without such stability, however, there would be no reason even to expect *distinct* manifestations of matter, to *identify* them, or to give them a name, let alone to start forming concepts or theories about them.

We are going to illustrate the phenomenology just described using a small simulation of a magnetic model-system. The system we shall be looking at is a so-called Ising ferromagnet. Macroscopic magnetic properties in such a system appear as average over microscopic magnetic moments attached to 'elementary magnets' called spins, each of them capable of two opposite orientations in space. These orientations can be thought of as parallel or anti-parallel to one of the crystalline axes. Model-systems of this kind have been demonstrated to capture magnetic properties of certain anisotropic crystals extremely well.

Denoting by $x_i(t) = \pm 1$ the two possible states of the i-th spins at time t, and by $\boldsymbol{x}(t)$ the configuration of *all* $x_i(t)$, one finds the macroscopic magnetisation of a system consisting of N such spins to be given by the *average*

$$S_N(\boldsymbol{x}(t)) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) \ . \tag{1}$$

In the model system considered here, a stochastic dynamics at the microscopic level is realised via a probabilistic time-evolution of the following form: Every spin experiences a 'local field'

$$u_i(t) = \sum_{i=1}^{N} J_{ij} x_j(t) \tag{2}$$

generated by all other spins; the interaction constants J_{ij} denote the relative size of the contribution of $x_j(t)$ to the local field $u_i(t)$ of the i-th spin. The spins determine their new state (in random order) following a probabilistic rule of the form

$$\mathrm{Prob}\Big\{x_i(t+\Delta t) = \pm 1\Big\} = \frac{1}{2}\Big[1 \pm \tanh(\beta u_i(t))\Big] \ . \tag{3}$$

The parameter β in this expression is inversely proportional to the absolute temperature T of the system; units can be chosen such that $\beta = 1/T$. The parameter β determines the degree of stochasticity of the dynamics: the larger β, (the lower the temperature), the larger the probability that $x_i(t + \Delta t)$ takes the value given by the direction (the sign) of the local field $u_i(t)$. One can easily convince oneself that positive interaction constants, $J_{ij} > 0$, encourage parallel orientation of spins, i.e. tendency to macroscopic ferromagnetic order.

[4] A diamond of 1 carat would already contain 100 times as many, viz. 10^{22} carbon atoms.

Independently of initial conditions, the stochastic dynamics (3) will approach a thermodynamic equilibrium state, described by a Gibbs-Boltzmann equilibrium distribution of micro-states

$$P(\boldsymbol{x}) = \frac{1}{Z_N} \exp\Big[- \beta H_N(\boldsymbol{x}) \Big] \tag{4}$$

corresponding to the 'energy function"

$$H_N(\boldsymbol{x}) = - \sum_{(i,j)} J_{ij} x_i x_j \ . \tag{5}$$

The double sum in this expression is over all possible pairs (i, j) of spins. In general, one

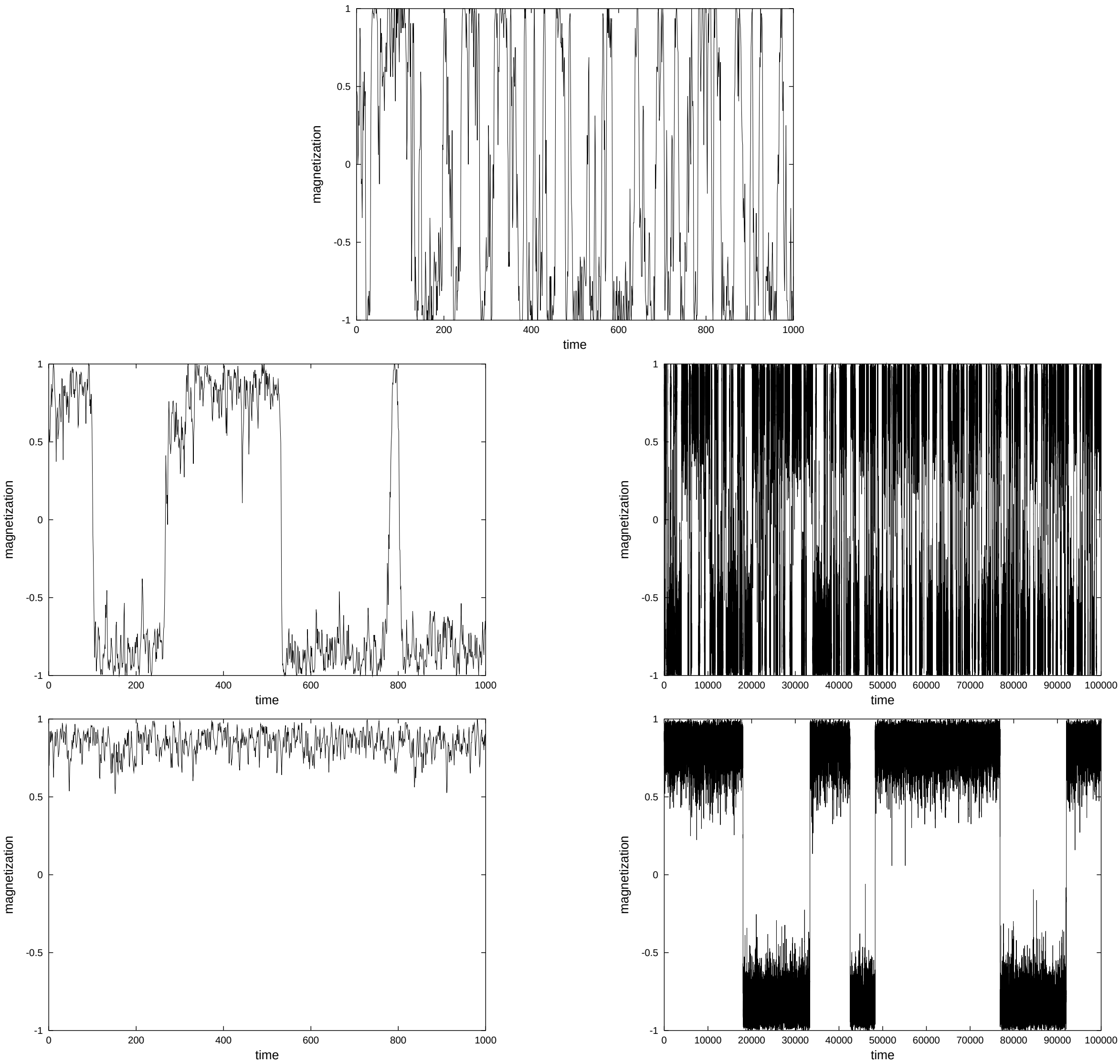

Figure 1: 1st panel: magnetisation of a system of $N = 8$ spins, monitored for 1000 time-steps; 2nd row: magnetisation of a system of $N = 32$ spins, monitored for 1000 time-steps (left panel) and 10^5 time-steps (right panel); 3rd row: magnetisation of a system of $N = 64$ spins, monitored for 1000 time-steps (left panel) and 10^5 time-steps (right panel). The temperature T in these simulations is chosen as $T = \frac{2}{3} T_c$, where T_c the temperature below which macroscopic systems of the type simulated would be magnetic.

expects the coupling stenghts J_{ij} to decrease as a function of the distance between spins i and j, and that they will tend to be negligibly small (possibly zero) at distances larger than a maximum range of the magnetic interaction.

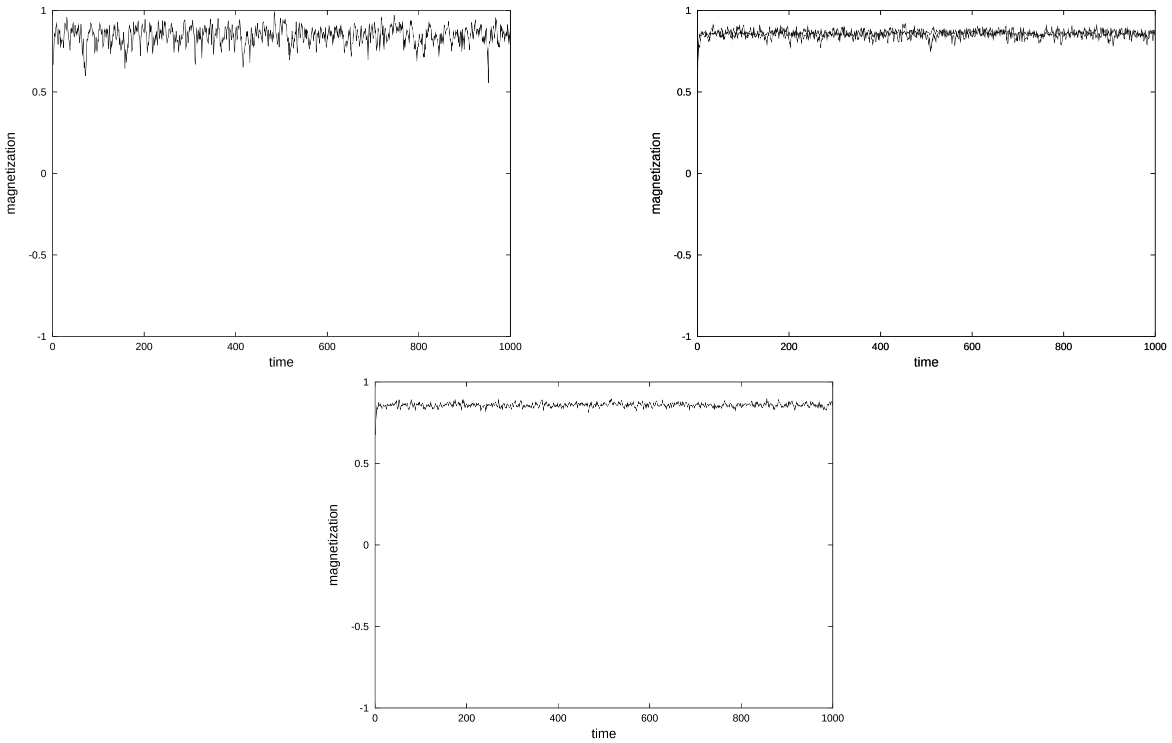

Figure 2: Magnetisation of systems containing $N = 128$, $N = 512$ and $N = 2048$ spins, monitored for 1000 time-steps.

Fig. 1 shows the magnetisation (1) as a function of time for various small systems. For the purposes of this figure, the magnetisation shown is already averaged over a time unit.[5]

The first panel of the figure demonstrates that a system consisting of 8 spins does not exhibit a stable value of its magnetisation. Increasing the number of spins by a factor 4 to 32, one observes in the left panel of the second row that the system appears to prefer values of its magnetisation around $S_N(\boldsymbol{x}(t)) \simeq \pm 0.8$. However, following the time-evolution of the magnetisation over a larger time span, which in real-time would correspond to approximately 10^{-7} seconds, as shown in the right panel of the second row, one notes that a stable or preferred value of the magnetisation is no longer discernible. If the system size is once more increased to 64 spins, a state with magnetisation fluctuating around 0.8 appears to be stable for a time span of around 10^{-9} seconds; yet sudden changes of the sign of the magnetisation still occur at irregular times, with time lags between switches of the order of several 10^{-8} seconds.

The experiment may be continued for larger and larger system sizes, as shown in Fig 2 for systems containing 128, 512 and 2048 spins. The fluctuations of the magnetisation become smaller as the system size is increased, and the time spans over which a stable non-switching magnetisation is observed increase with system size as shown for the smaller systems in Fig

[5]The time unit in these simulations is given by the time span during which every single spin has on average once been selected for an update of its state according to (3). It is this unit of time which is comparable for systems of different sizes [5]; it would correspond to a time-span of approximately 10^{-12} seconds in conventional units.

1. Indeed, the system containing $N = 2048$ spins may already be expected to exhibit a magnetisation which remains stable for times of the order of a minute. However, the system would still be useless as a magnetic material to build a compass with, as its magnetisation would switch sign randomly at times in the one minute range. Only in the infinite system limit would a system exhibit a constant non-fluctuating magnetisation, and only in this limit would one therefore, strictly speaking, be permitted to talk of a system with a given value of its magnetisation.

The present example system was set up in a way that states with magnetisations $S_N \simeq \pm 0.8$ would be its only (two) distinct macroscopic manifestations. Transitions between them are observed at irregular time intervals as long as the system is finite. This is the macroscopic manifestation of ergodicity. The time span over which a given macroscopic manifestation is stable will diverge — ergodicity can be broken — only in the infinitely large system.

3 Limit Theorems and Description on Large Scales

Large numbers are according to our reasoning a prerequisite for stability of macroscopic material properties, and only in the limit of large numbers we may expect that macroscopic properties of matter are also non-fluctuating. Early formulations of equations of state of macroscopic systems which postulate deterministic functional relations, e.g. between temperature, density and pressure of a gas, or temperature, magnetisation and magnetic field in magnetic systems can therefore claim strict validity only in in the infinite system limit. They are thus seen to presuppose this limit, though in most cases, it seems, implicitly.

From a mathematical perspective, there are — as already stressed by Khinchin in his classical treatise on the mathematical foundations of statistical mechanics [6] — two limit theorems of probability theory, which are of particular relevance for the phenomena just described: (i) the law of large numbers, according to which a sum of N identically distributed random variables of the form (1) will in the limit $N \to \infty$ converge to the expectation $\mu = \langle x_k \rangle$ of the x_k, assuming that the latter is finite; (ii) the central limit theorem according to which the die distribution of deviations from the expectation, i.e., the distribution of $S_N - \mu$ for independent identically distributed random variables will in the limit of large N converge to a Gaussian of variance σ^2/N, where $\sigma^2 = \langle x_k^2 \rangle - \langle x_k \rangle^2$ denotes the variance of the x_k.[6]

The central limit theorem in particular implies that fluctuations of macroscopic quantities of the form (1) will in the limit of large N typically decrease as $1/\sqrt{N}$ with system size N. This result, initially formulated for independent random variables may be extended to so-called weakly dependent variables. Considering the squared deviation $(S_N - \langle S_N \rangle)^2$, one obtains its expectation value

$$\Big\langle (S_N - \langle S_N \rangle)^2 \Big\rangle = \frac{1}{N^2} \sum_{k,\ell=1}^{N} C_{k,\ell} = \frac{1}{N^2} \sum_{k,\ell=1}^{N} \Big\langle (x_k - \langle x_k \rangle)(x_\ell - \langle x_\ell \rangle) \Big\rangle \tag{6}$$

and the desired extension would hold for all systems, for which the correlations $C_{k,\ell}$ are decreasing sufficiently rapidly with "distance" $|k - \ell|$ to ensure that the sums $\sum_{\ell=1}^{\infty} C_{k,\ell}$ are finite for all k.

The relation between the above-mentioned limit theorems and the description of stochastic systems at large scales are of particular interest for our investigation, a connection that was first pointed out by Jona-Lasinio [8].[7] The concept of large-scale description has been

[6] For precise formulations of conditions and proofs, see Feller [7].

[7] On this, see also Batterman [9], who referred to the relation on several occasions in the context of debates on reductionism.

particularly influential in the context of the renormalisation group approach which has led to the our current and generally accepted understanding of critical phenomena.[8]

To discuss this relation, let us return to independent random variables and, generalising Eq. (1), consider sums of random variables of the form

$$S_N(\boldsymbol{x}) = \frac{1}{N^{1/\alpha}} \sum_{k=1}^{N} x_k \ . \tag{7}$$

The parameter α fixes the power of system size N by which the sum must be rescaled in order to achieve interesting, i.e., non-trivial results. Clearly, if α is too small (for the type of random variables considered), then the power of N will be large enough to make the sum almost surely vanish, $S_N \to 0$, in the large N limit. Conversely, if α is too large, then the power of N will be small enough to make the sum almost surely diverge, $S_N \to \pm\infty$, as N becomes large. We shall in what follows restrict our attention to the two important cases $\alpha = 1$ — appropriate for sums of random variables of non-zero mean — and $\alpha = 2$ — relevant for sums of random variables of zero mean and finite variance. For these two cases, we shall recover the propositions of the law of large numbers ($\alpha = 1$) and of the central limit theorems ($\alpha = 2$) as properties of *large-scale* descriptions of (7).

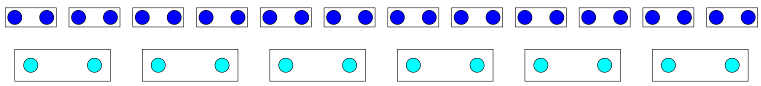

Figure 3: Repeated enlargement of scale: the present example begins with a system of 24 random variables symbolised by the dots in the first row. These are combined in pairs, as indicated by frames surrounding two neighbouring dots. Each such pair generates a renormalised variable, indicated by 12 dots of lighter shade in the second row of the figure. Two of those are combined in the next step as indicated by frames around pairs of renormalised variables, thereby starting the iteration of the renormalisation procedure.

To this end we imagine the x_k to be arranged on a linear chain. The sum (7) may now be reorganised by (i) combining neighbouring pairs of the original variables and computing their averages $\overline{x}_k$, yielding $N' = N/2$ of such local averages, and by (ii) appropriately rescaling these averages so as to obtain renormalised random variables $x'_k = 2^\mu \overline{x}_k$, and by expressing the original sum (7) in terms of a corresponding sum of the renormalised variables, formally

$$\overline{x}_k = \frac{x_{2k-1} + x_{2k}}{2} \ , \qquad x'_k = 2^\mu \overline{x}_k \ , \qquad N' = N/2 \ , \tag{8}$$

so that

$$S_N(\boldsymbol{x}) = 2^{(1-1/\alpha-\mu)} S_{N'}(\boldsymbol{x}') \ . \tag{9}$$

One may compare this form of "renormalisation" — the combination of local averaging and rescaling — to the effect achieved by combining an initial reduction of the magnification of a microscope (to the effect that only locally averaged features can be resolved) with an ensuing change of the contrast of the image it produces.

By choosing the rescaling parameter μ such that $\mu = 1 - 1/\alpha$, one ensures that the sum (7) remains invariant under renormalisation,

$$S_N(\boldsymbol{x}) = S_{N'}(\boldsymbol{x}') \ . \tag{10}$$

[8]The notion of critical phenomena refers to a set of anomalies (non-analyticities) of thermodynamic functions in the vicinity of continuous, second order phase transitions. A lucid exposition is given by Fisher [11].

The renormalisation procedure may therefore be iterated, as depicted in Fig. 3: $x_k \to x'_k \to x''_k \to \ldots$, and one would obtain the corresponding identity of sums expressed in terms of repeatedly renormalised variables,

$$S_N(\boldsymbol{x}) = S_{N'}(\boldsymbol{x}') = S_{N''}(\boldsymbol{x}'') = \ldots \tag{11}$$

The statistical properties of the renormalised variables x'_k will, in general be different from (though, of course, dependent on) those of the original variables x_k, and by the same token will the statistical properties of the doubly renormalised variables x''_k be different from those of the x'_k, and so on. However, one expects that statistical properties of variables will after sufficiently many renormalisation steps, i.e., at large scale, eventually become independent of the microscopic details and of the scale considered, thereby becoming largely independent of the statistical properties of the original variables x_k, and invariant under further renormalisation. This is indeed what happens under fairly general conditions.

It turns out that for sums of random variables x_k with non-zero average $\mu = \langle x_k \rangle$ the statement of the law of large numbers is recovered. To achieve asymptotic invariance under repeated renormalisation, one has to choose $\alpha = 1$ in (7), in which case one finds that the repeatedly renormalised variables $x_k^{''''\cdots}$ converge under repeated renormalisation to the average of the x_k, which is thereby seen to coincide with the large N limit of the S_N.

If sums of random variables of zero mean (but finite variance) are considered, the adequate scaling is given by $\alpha = 2$. In this case the repeatedly renormalised variables $x_k^{''''\cdots}$, and thereby the S_N, are asymptotically normally distributed with variance σ^2 of the original variables x_k, even if these were not themselves normally distributed. The interested reader will find details of the mathematical reasoning underlying these results in an appendix at the end of this paper.

Let us not fail to mention that other stable distributions of the repeatedly renormalised variables $x_k^{''''\cdots}$, thus of the S_N — the so-called Lévy α-stable distributions [7] — may be obtained by considering sums random variables of infinite variance. Although such distributions have recently attracted some attention in the connection with the description of complex dynamical systems, such as turbulence or financial markets, they are of lesser importance for the description of thermodynamic systems in equilibrium, and we shall therefore not consider these any further in what follows..

4 Interacting systems and the renormalisation group

For the purpose of describing macroscopic systems the concept of large-scale descriptions of a system, used above to elucidate the two main limit theorems of mathematical statistics, needs to be generalised to interacting, thus *correlated* or dependent random variables. Such a generalisation was formulated at the beginning of the 1970s as renormalisation group approach to interacting systems.

Starting point of this approach is the Boltzmann-Gibbs equilibrium distribution of microscopic degrees of freedom taking the form (4). The idea of the renormalisation group approach to condensed matter systems is perhaps best explained in terms of the normalisation constant Z_N appearing in (4), the so-called partition function. It is related to the dimensionless free energy $\overline{f}_N$ of the system via $Z_N = \mathrm{e}^{-N\overline{f}_N}$ and thereby to its thermodynamic functions and properties.[9] To this end the partition function in (4) is written in the form

$$Z_N = Z_N(\boldsymbol{K}) = \sum_{\boldsymbol{x}} \mathrm{e}^{-\overline{H}_N(\boldsymbol{x};\boldsymbol{K})} \ , \tag{12}$$

in which $\overline{H}_N(\boldsymbol{x};\boldsymbol{K})$ denotes the dimensionless energy function of the system, i.e., the conventional energy function multiplied by the inverse temperature β, while $\boldsymbol{K}$ stands for the collection of all coupling constants in H_N (multiplied by β). These may include two-particle couplings as in (5), but also single-particle couplings as well as a diverse collection of many-particle couplings. Renormalisation investigates, how the formal representation of the partition function changes, when it is no longer interpreted as a sum over all micro-states of the original variables, but as a sum over micro-states of renormalised variables, the latter defined as suitably rescaled local averages of the original variables in complete analogy to the case of independent random variables.

In contrast to the case of independent variables, geometric neighbourhood relations play a crucial role for interacting systems, and are determined by the physics of the problem. E.g., for degrees of freedom arranged on a d-dimensional (hyper)-cubic lattice, one could average over the b^d degrees of freedom contained in a (hyper)-cube B_k of side-length b to define locally averaged variables, as illustrated in Fig. 4 for $d = 2$ and $b = 2$, which are then rescaled by a suitable factor b^μ in complete analogy to the case of independent random variables discussed above,

$$\overline{x}_k = b^{-d} \sum_{i \in B_k} x_i \ , \qquad x'_k = b^\mu \overline{x}_k \ . \qquad N' = N/b^d \tag{13}$$

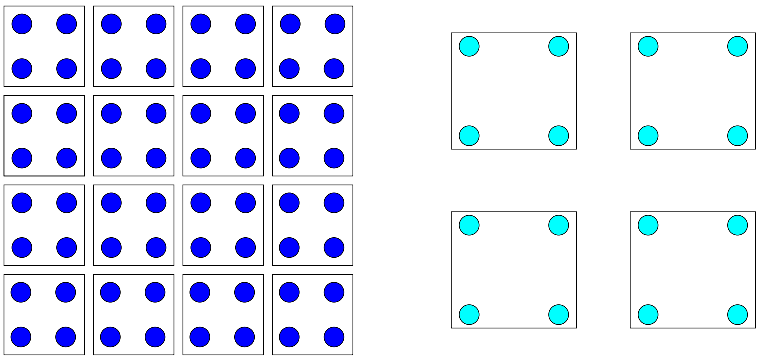

Figure 4: Iterated coarsening of scale

The partition sum on the coarser scale is then evaluated by first summing over all micro-states of the renormalised variables $\boldsymbol{x}'$ and for each of them over all configurations $\boldsymbol{x}$ compatible with the given $\boldsymbol{x}'$, formally

$$Z_N(\boldsymbol{K}) = \sum_{\boldsymbol{x}'} \left[\sum_{\boldsymbol{x}} P(\boldsymbol{x}', \boldsymbol{x}) \mathrm{e}^{-\overline{H}_N(\boldsymbol{x};\boldsymbol{K})} \right] \equiv \sum_{\boldsymbol{x}'} \mathrm{e}^{-\overline{H}_{N'}(\boldsymbol{x}';\boldsymbol{K}')} = Z_{N'}(\boldsymbol{K}') \tag{14}$$

where $P(\boldsymbol{x}', \boldsymbol{x}) \geq 0$ is constructed in such a way that $P(\boldsymbol{x}', \boldsymbol{x}) = 0$, if $\boldsymbol{x}$ is incompatible with $\boldsymbol{x}'$, whereas $P(\boldsymbol{x}', \boldsymbol{x}) = \rho(\boldsymbol{x}')$, if $\boldsymbol{x}$ is compatible with $\boldsymbol{x}'$, and $\rho(\boldsymbol{x}')$ normalised such that

[9] The dimensionless free energy is just the product of the standard free energy and the inverse temperature β. At a formal level, the partition function is closely related to the characteristic function of a (set of) random variables, in terms of which we analysed the idea of large-scale descriptions for sums of independent random variables in appendix A.

$\sum_{\boldsymbol{x}'} P(\boldsymbol{x}', \boldsymbol{x}) = \sum_{\boldsymbol{x}'} \rho(\boldsymbol{x}') = 1 \;\forall\; \boldsymbol{x}$. The result is interpreted as the partition function corresponding to a system of $N' = b^{-d}N$ renormalised variables, corresponding to a dimensionless energy function $\overline{H}_{N'}$ of the same format as the original one, albeit with renormalised coupling constants $\boldsymbol{K} \to \boldsymbol{K}'$, as expressed in (14). The distance between neighbouring renormalised degrees of freedom is larger by a factor b than that of the original variables. Through an ensuing rescaling of all lengths $\ell \to \ell/b$ one restores the original distance between the degrees of freedom, and completes the renormalisation group transformation as a mapping between systems of the same format.

As in the previously discussed case of independent random variables, the renormalisation group transformation may be iterated and thus creates not only a sequence of repeatedly renormalised variables, but also a corresponding sequence of repeatedly renormalised couplings

$$\boldsymbol{K} \to \boldsymbol{K}' \to \boldsymbol{K}'' \to \boldsymbol{K}''' \to \dots \; . \tag{15}$$

As indicated in Fig. 5, this sequence may be visualised as a renormalisation group *'flow'* in the space of couplings.

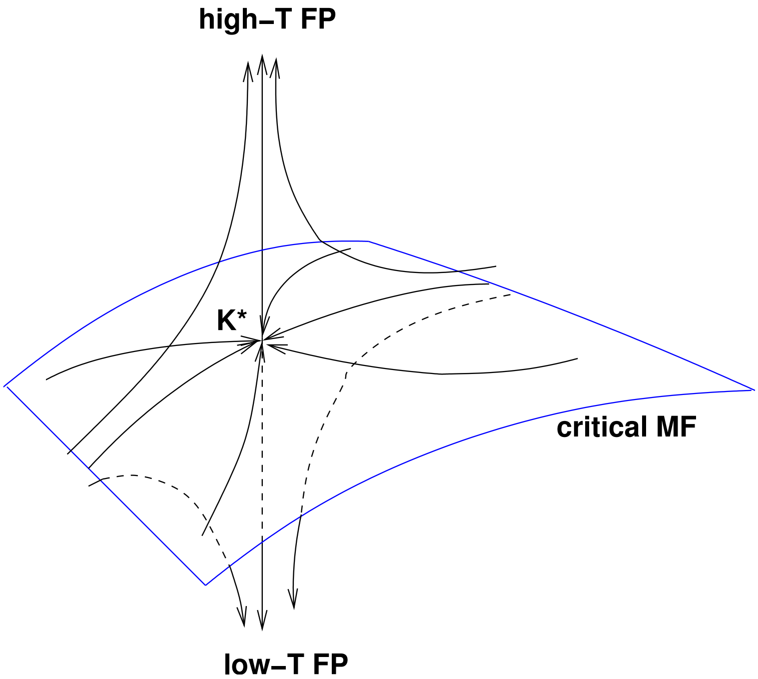

Figure 5: Renormalisation group flow in the space of couplings

The renormalisation transformation entails a transformation of (dimensonless) free energies $\overline{f}_N(\boldsymbol{K}) = -N^{-1} \ln Z_N$ of the form

$$\overline{f}_N(\boldsymbol{K}) = b^{-d}\overline{f}_{N'}(\boldsymbol{K}') \; . \tag{16}$$

For the present discussion, however, the corresponding transformation of the so-called correlation length ξ which describes the distance over which the degrees of freedom in the system are statistically correlated, is of even greater interest. As a consequence of the rescaling of all lengths, $\ell \to \ell/b$, in the final step of the renormalisation transformation, one obtains

$$\xi_N(\boldsymbol{K}) = b\,\xi_{N'}(\boldsymbol{K}') \; . \tag{17}$$

Repeated renormalisation amounts to a description of the system on larger and larger length scales. The expectation that such a description would on a sufficiently large scale eventually become independent of the scale actually chosen would correspond to the finding that the renormalisation group flow would typically approach a fix-point: $\boldsymbol{K} \to \boldsymbol{K}' \to \boldsymbol{K}'' \to \ldots \to \boldsymbol{K}^*$.

The existence of fixed points is of particular significance in the limit of infinitely large system size $N = N' = \infty$, as in this limit Eq. (17), $\xi_\infty(\boldsymbol{K}^*) = b\xi_\infty(\boldsymbol{K}^*)$ will for $b \neq 1$ *only* allow for the two possibilities

$$\xi_\infty(\boldsymbol{K}^*) = 0 \qquad \text{or} \qquad \xi_\infty(\boldsymbol{K}^*) = \infty \ . \tag{18}$$

The first either corresponds to a so-called high-temperature fixed point, or to a low-temperature fixed point. The second possibility with infinite correlation length corresponds to a so-called critical point describing a continuous, or second order phase transition. In order to realise the second possibility, the parameters of the system must be adjusted in such a way that they come to lie precisely on the critical manifold in the space of parameters (see Fig. 5)[10]

For the purpose of the present discussion, however, the phenomenology close to off-critical high and low-temperature fixed points is of even greater importance, as it entails that degrees of freedom are virtually *uncorrelated* on large scales, and that therefore the description of non-critical systems within the framework of the two limit theorems for independent variables discussed earlier is *entirely adequate*, despite the correlations over small distances created by interactions between the original microscopic degrees of freedom.

5 The thermodynamic limit of infinite system size

We are ready for a first summary: only in the thermodynamic limit of infinite system size $N \to \infty$ will macroscopic systems exhibit stable and non-fluctuating thermodynamic properties; only in this limit can we expect that deterministic equations of state exist which describe relations between different thermodynamic properties as well as the manner in which these depend on external parameters such as pressure, temperature oder electromagnetic fields.

It is quite remarkable that equations of state for macroscopic systems are usually also *simple* in the sense that they require just very few *relevant* quantities for a complete thermodynamic description of such systems (such as pressure, temperature and density in the case of gasses and liquids) — a fact that is rarely ever emphasised as it has become so commonplace, although it would in principle seem to require explanation. This is not the place to supply such an explanation; suffice it to mention that the relevant quantities are just the ones related to parameters that are relevant for a large-scale description in the renormalisation group sense.[11] In order that the remaining, in a technical sense irrelevant variables cannot influence thermodynamic properties, it is required that a sufficiently coarse level of description of a system does *exist* at all; this requires, strictly speaking, once more that the thermodynamic limit is taken.

In less technical terms, the observation that thermodynamical properties of macroscopic systems will typically be *independent* of the shape of a sample would have to be mentioned in this context[12] — this, being related to the fact that the fraction of constituents that feel the

[10]Experience shows that this typically requires to fix only a few parameters within the high-dimensional space of parameters to their critical values. For conventional continuous phase transitions these are normally two — temperature and pressure in the case of gasses, temperature and magnetic field in magnetic systems, and so forth. The fact that all systems on the critical manifold are controlled by the same fixed point does in itself have the remarkable consequence that there exist large classes of microscopically very diverse systems, the so-called universality classes, which exhibit essentially the same behaviour at their respective critical points [10, 11].

[11]These are the couplings (and their associated fields), whose distance from critical fixed points in the space of couplings is increased under renormalisation.

[12]Well-understood exceptions exist for systems with truly long-range (e.g., dipolar) interactions.

influence of boundaries and their shapes will be negligibly small only in the thermodynamic limit. This is frequently exploited in mathematical analyses of such systems by choosing them to have periodic boundary conditions (by closing a line to a circle, a plane to a torus, or — not easily imagined in our three-dimensional world — a prism to a hyper-torus). Such systems will then no longer exhibit *any* boundaries, and all locations within such systems become strictly equivalent, thereby often allowing considerable simplifications in an ensuing mathematical analysis. In the context of the limit theorems of statistics, for instance, such an elimination of boundaries would strictly speaking be *required* in order to consider the degrees of freedom as equivalent, hence identically distributed. Clearly, every boundary in the system would contradict equivalence assumptions on degrees of freedom, as those at or close to a boundary would clearly have to behave differently from those deep in the bulk.

Real systems are, of course, always finite. We are thus always dealing with the potentially infinite in the Aristotelian sense. The decisive aspect for physics and for our ability to capture the essential physical properties of macroscopic systems in terms of reasonably simple and manageable theories is that such systems are invariably composed of huge numbers of atomic or molecular constituents, numbers so large in fact that they are for practical purposes, i.e. for determining physical properties of these systems, indistinguishable from infinity.

We saw that this aspect is responsible for regularity and stability of macroscopic properties of such systems, as well as for the structural simplicity of relations that exist between different physical properties, and for these reasons constitutive for our ability to formulate theories for them.

6 Neural information processing

Having discussed the constitutive role of large numbers for our ability to create theories about macroscopic systems, we shall now move on to an argument according to which — given our current understanding of information processing in neural systems — the prominent role of large numbers for our mere ability of constructing theories appears to reach *much* further than our initial discussion within the domain of theorising about macroscopic systems would suggest.

The argument links up with our discussion of the empirical foundations of our main hypothesis in Sect 2, in which we emphasise the role of large numbers for the emergence of regularity and stability in stochastic systems. Whereas we were concerned with the stability of observed (and observable) phenomena, we are here considering in particular the *processing of* observed regularities by an information processing apparatus consisting of a collection of neurons.

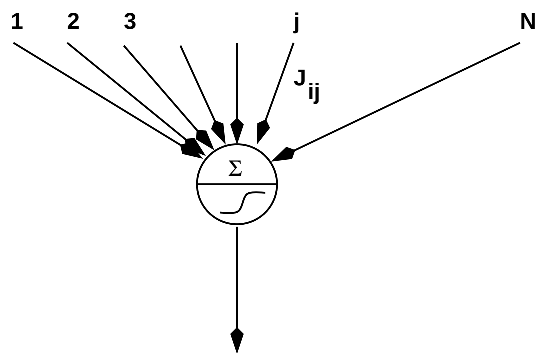

Figure 6: Information processing by a single neuron

Neurons may be schematically described as threshold-elements. They experience a post-synaptic potential,

$$u_i = \sum_{j=1}^{N_i} J_{ij} x_j \ , \tag{19}$$

a sum input signals x_j transmitted to them from other neurons or directly from cells of sensory organs via synaptic connections, and in transmission weighted by the strengths J_{ij} of these connections (see Fig. 6). The input-signal $x_j = x_j(t)$ specifies, whether or not an action potential was received through input channel j at the specified time t. We denote by N_i the number of input-channels of neuron i. The response of the neuron depends on the value of its post-synaptic potential. If this value remains below a given threshold ϑ_i, the neuron will remain quiescent, otherwise it will fire an action potential with a probability

$$\nu_i = g(u_i - \vartheta_i) \tag{20}$$

that depends on the comparison of the post-synaptic potential u_i and a threshold ϑ_i. The action potential will then travel along the neuron's axon and its ramifications, and be relayed to N_i^o posterior neurons. The firing probability ν_i is a monotone increasing function g of the distance of post-synaptic potential and threshold, which is close to zero below threshold, and rises to appreciable values only above threshold.[13]

Now, it is important to realise that both the mechanisms, underlying the creation input signals by sensory cells, as well as those underlying synaptic transmission and the generation of action potentials, are fundamentally stochastic in nature (see e.g. [12]). It is for this reason that the collection of input signals $\{x_j; j = 1, N_i\}$ of a given neuron i can — even for a given, unchanging state of the external world — at best be characterised by its statistical properties. Post-synaptic potentials would thus have the character of randomly weighted sums of random variables with statistical properties that are, other things being equal, determined by states of the external world, in as much as, indeed we are considering the neural representation of such states. In order to exhibit a certain degree of reproducibility in the case of reproducible states of the external world,[14] these sums would have to involve large sums of input signals. Due to the stochastic nature of the processes underlying the generation of action potentials and of synaptic transmission, a reliable representation of states of the external world — and for the same reasons a reliable representation of states of the 'internal world' of an organism — can only be guaranteed by the fact that such representations invariably involve *large numbers of neurons*.[15] Without claiming that that the processes leading to conscious perception of states of the external or the internal world be even approximately understood,[16] it would seem fairly safe to assume, that these processes will at *all* levels require participation of large neural populations — at least, that is, if it is conceded that they cannot be described *without* a concurrent dynamics of brain states.[17]

We thus feel justified to conclude that large numbers are not only a prerequisite for our ability to theorise about macroscopic systems by being constitutive for the stability of their

[13] The present description is highly schematised and simplified; taking more details into account would if anything, however, only further enforce the argument presented in what follows.

[14] It should be obvious that this is a minimal requirement to survive for a higher organism which is able to, and depends on appropriately reacting to its environment.

[15] Neuro-anatomical data do at least not explicitly stand in contradistinction to this observation: the human brain contains roughly 10^{11} neurons; each of them typically receives input signals from 10^4 other neurons and relays its output signals typically to 10^4 other neurons.

[16] See [13] for an extensive discussion of this issue.

[17] The wording here is carefully chosen to acknowledge the fact that the question is still open, whether the processes leading to conscious perception can be fully described in terms of brain states. There exists, in fact, broad spectrum of positions on this question [13]. However, it now appears to be fairly generally accepted that there is at least a neural correlate of these processes.

properties, but that they are also of fundamental importance for (neural) perception and and representation of the world around and inside us.

7 Summary

We demonstrated that large numbers are essential for the emergence of statistical regularities in macroscopic systems. Without such regularity, however, projects aiming at theoretical understanding of natural phenomena could not have been feasible or successful. From a mathematical point of view the propositions of two limit theorems, the law of large numbers and the central limit theorem were seen to be of crucial importance for the emergence of such regularities, and we demonstrated that these are intimately related to the idea of large scale descriptions of such systems and thereby with the renormalisation group concept. The limit theorems are strictly valid only in the limit of infinite system size. The fact that real finite systems partake of the simplifications resulting from descriptions in terms of limit theorems is due circumstance that these systems are composed of so many particles as to make their statistical properties for all practical purposes indistinguishable from those of truly infinite systems. We saw that this fact is also responsible for the fact that equations of state describing relations between macroscopic properties of such systems only involve a small set of variables for a full thermodynamic characterisation.

We have strong reasons to suspect that the constitutive role of large numbers for our ability to construct theories extends far beyond the domain of theories for macroscopic systems which we used as the starting point of our discussion.

A Renormalisation and characteristic functions

The renormalisation group transformation for the case of sums of independent random variables is bes investigated in terms of their *characteristic functions*.

The characteristic function of a random variable X is defined as the Fourier transform of its probability density p_X,[18]

$$\varphi_X(k) = \left\langle \mathrm{e}^{ikX} \right\rangle = \int \mathrm{d}x \; p_X(x) \; \mathrm{e}^{ikx} \; . \tag{21}$$

Characteristic functions are important tools in probability. Among other things, they can be used to express moments of a random variable in compact form via differentiation,

$$\left(-i\frac{\mathrm{d}}{\mathrm{d}k}\right)^n \varphi_X(k)\bigg|_{k=0} = \langle X^n \rangle = \int \mathrm{d}x \; p_X(x) \; x^n \; . \tag{22}$$

The second important property needed here is that the characteristic function of a sum $X+Y$ of two independent random variables X and Y is given by the product of their characteristic functions. For, denoting by p_X and p_Y the probability densities corresponding to the two variables, one finds

$$\varphi_{X+Y}(k) = \int \mathrm{d}x\mathrm{d}y \; p_X(x)p_Y(y) \; \mathrm{e}^{ik(x+y)} = \varphi_X(k)\varphi_Y(k) \; . \tag{23}$$

We investigate properties of (7),

$$S_N(\boldsymbol{x}) = \frac{1}{N^{1/\alpha}} \sum_{k=1}^{N} x_k$$

[18] In this appendix, we follow the mathematical convention to distinguish in notation between a random variable X and its realisation x.

under renormalisation, using its characteristic function to be denoted by ϕ_N. Let φ_1 denote the characteristic function of the original variables, φ_2 that of the renormalised variables x'_k (constructed from sums of two of the original variables), and more generally, let φ_{2^ℓ} denote the characteristic function of the ℓ-fold renormalised variables, constructed from sums involving 2^ℓ original variables. We then get

$$\begin{aligned} \phi_N(k) &:= \left\langle \mathrm{e}^{ikS_N} \right\rangle = \varphi_1\left(\frac{k}{N^{1/\alpha}}\right)^N = \varphi_2\left(\frac{k}{(N/2)^{1/\alpha}}\right)^{N/2} \\ &= \varphi_4\left(\frac{k}{(N/4)^{1/\alpha}}\right)^{N/4} = \ldots = \varphi_{2^\ell}\left(\frac{k}{(N/2^\ell)^{1/\alpha}}\right)^{N/2^\ell} \end{aligned} \tag{24}$$

Assuming that multiply renormalised variables will acquire asymptotically stable statistical properties, i.e. statistical properties that remain invariant under further renormalisation, the φ_{2^ℓ} would have to converge to a limiting function φ^*,

$$\varphi_{2^\ell} \to \varphi^* \qquad , \ \ell \to \infty \ . \tag{25}$$

This limiting function φ^* would have to satisfy a functional *self-consistency* relation of the form

$$\varphi^*(k) = \varphi^*(2^{1/\alpha}k)^{1/2} \tag{26}$$

which follows from (24), using the invariance of S_N under repeated renormalisation.

The solutions of this self-consistency relation for $\alpha = 1$ and $\alpha = 2$ are seen to be given by

$$\varphi^*(k) \equiv \lim_{N\to\infty} \phi_N(k) = \exp(-ck^\alpha) \ , \tag{27}$$

and it can be shown that these solutions are unique in the space of characteristic functions for random variables having finite moments.

One identifies the characteristic function of a non-fluctuating (i.e. constant) random variable with $c = -i\langle X\rangle = -i\mu$ for $\alpha = 1$, and the characteristic function of a Gaussian normal density with $c = \frac{1}{2}\sigma^2$ for $\alpha = 1/2$, and thereby verifies the statements of the two limit theorems.

One can also show that the convergence (25) is realised for a very broad spectrum of distributions for the microscopic variables, both for $\alpha = 1$ (the law of large numbers), and for $\alpha = 1/2$ (the central limit theorem). For $\alpha = 1$, there is a "marginal direction" in the infinite-dimensional space of possible perturbations of the invariant characteristic function (corresponding to a change of the expectation value of the random quantities being summed), which doesn't change its distance to the invariant function $\varphi^*(k)$ under renormalisation. All other perturbations are irrelevant in the sense that their distance from the invariant characteristic function will diminish under repeated renormalisation. For $\alpha = 2$ there is one "relevant direction" in the space of possible perturbations, in which perturbations of the invariant characteristic function will be amplified under repeated renormalisation (it corresponds to introducing a non-zero mean of the random variables being added), and a marginal direction that corresponds to changing the variance of the original variables. All other perturbations are irrelevant and will be diminished under renormalisation. The intersted reader will find a formal verification of these statements in the following Appendix B.

B Linear stability analysis

Statements about the stability of invariant characteristic functions under various perturbations are proved by looking at the linearisation of the renormalisation group transformation in the vicinity of the invariant distribution. We shall adhere to the formulation using characteristic functions which makes the full analysis somewhat easier than the one in terms of probability densities used in [14].

Let R_α denote the renormalisation transformation of a characteristic function for the scaling exponent α. From (24), we see that its action on a characteristic function φ is defined as

$$R_\alpha[\varphi](2^{1/\alpha}k) = \varphi(k)^2 \tag{28}$$

Assuming $\varphi = \varphi^* + h$, where h is a small perturbation of the invariant characteristic function, whe have

$$R_\alpha[\varphi^* + h](2^{1/\alpha}k) = (\varphi^*(k) + h(k))^2 \simeq \varphi^*(k)^2 + 2\varphi^*(k)h(k)\ , \tag{29}$$

where the expansion on the r.h.s. has been carried to first order in the small perturbation h. Using an expansion of the transformation R_α in the vicinity of φ^*, and denoting by $D_\alpha = D_\alpha[\varphi^*]$ the operator of the linearised transformation in the vicinity of φ^* on the l.h.s, one has $R_\alpha[\varphi^* + h] \simeq R_\alpha[\phi^*] + D_\alpha h$ to linear order in h, thus

$$R_\alpha[\varphi^*](2^{1/\alpha}k) + D_\alpha h(2^{1/\alpha}k) \simeq \varphi^*(k)^2 + 2\varphi^*(k)h(k)\ . \tag{30}$$

By the invariance of φ^* under R_α, we get

$$D_\alpha h(2^{1/\alpha}k) = 2\varphi^*(k)h(k) \tag{31}$$

to linear order. The stability of the invariant characteristic function is then determined by the spectrum of D_α, found by solving the eigenvalue problem

$$D_\alpha h(2^{1/\alpha}k) = 2\varphi^*(k)h(k) = \lambda h(2^{1/\alpha}k)\ . \tag{32}$$

In the case where $\alpha = 1$ we have $\varphi^*(k) = \mathrm{e}^{-i\mu k}$, and the eigenvalue equation reads

$$2\mathrm{e}^{-i\mu k}h(k) = \lambda h(2k)\ . \tag{33}$$

To solve it we introduce f by writing

$$h(k) = \mathrm{e}^{-i\mu k}f(k)$$

which transforms the eigenvalue equation into

$$2f(k) = \lambda f(2k)\ . \tag{34}$$

Clearly this equation is solved by homogeneous functions:

$$2f_n(k) = \lambda_n f_n(2k)\ , \qquad f_n(k) = \frac{(ikx_0)^n}{n!}, \qquad \lambda_n = 2^{1-n}\ . \tag{35}$$

In order for $\varphi^* + h_n$ with $h_n(k) = \mathrm{e}^{-i\mu k}f_n(k)$ to be a characteristic function (of a system with finite moments), we must have $n \geq 1$, so that $\lambda_1 = 1$ (the corresponing perturbation being marginal), and $\lambda_n < 1$ (the corresponding perturbations thus being irrelevant) for all $n > 1$. The marginal perturbation amounts to changing the mean of the random variable to $\mu + x_0$, as mentioned earlier.

In the case where $\alpha = \frac{1}{2}$ we have $\varphi^*(k) = \mathrm{e}^{-\frac{1}{2}\sigma^2k^2}$, so the eigenvalue equation is

$$2\mathrm{e}^{-\frac{1}{2}\sigma^2k^2}h(k) = \lambda h(\sqrt{2}k)\ . \tag{36}$$

To solve it we now define f by writing

$$h(k) = \mathrm{e}^{-\frac{1}{2}\sigma^2k^2}f(k)$$

which transforms the eigenvalue equation into

$$2f(k) = \lambda f(\sqrt{2}k) \ . \tag{37}$$

Clearly this equation, too, is solved by homogeneous functions:

$$2f_n(k) = \lambda_n f_n(\sqrt{2}k) \ , \qquad f_n(k) = \frac{(ikx_0)^n}{n!}, \qquad \lambda_n = 2^{1-\frac{n}{2}} \ . \tag{38}$$

Once more, for $\varphi^* + h_n$ with $h_n(k) = \mathrm{e}^{-\frac{1}{2}\sigma^2 k^2} f_n(k)$ to be a characteristic function (of a system with finite moments), we must have $n \geq 1$, so that $\lambda_1 = 2^{\frac{1}{2}}$ (the corresponing perturbation being relevant), $\lambda_2 = 1$ (the corresponing perturbation being marginal), and $\lambda_n < 1$ for all $n > 2$ (the corresponding perturbations thus being irrelevant). In the present case, the relevant perturbation amounts to introducing a nonzero mean x_0 of the original random variables, while the marginal perturbation changes the variance to $\sigma^2 + x_0^2$, as mentioned earlier. All other perturbations change higher order cumulants of the random variables considered and are irrelevant.

References

[1] O. Neugebauer, *The Exact Sciences in Antiquity*, (Harper, New York, 1962).

[2] B. D'Espagnat, *Conceptual Foundations of Quantum Mechanics*, 2nd Edition, (W. A. Benjamin, Reading Mass., 1976).

[3] I. Kant, *Kritik der reinen Vernunft*, Erste Auflage (A), Werkausgabe Band III, Willhelm Weischedel (Hrsg.), p A100; Nachdruck im Suhrkamp Taschenbuch, stw **55**, (Suhrkamp, Frankfurt, 1974), p 163.;

[4] I. Kant, *Critique of Pure Reason*, English translation: N. Kemp Smith (1929), published as e-text, URL: `http://www.hkbu.edu.hk/~ppp/cpr/toc.html`

[5] K. Binder and D. Stauffer, "A Simple Introduction to Monte Carlo Simulation and Some Specialised Topics" , in: *Applications of the Monte Carlo Method in Statistical Physics*, 2nd edition, K. Binder (Ed.), (Springer, Berlin, Heidelberg, 1987), pp 1–36.

[6] A.I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, (Dover, New York, 1949).

[7] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd Edition, Vols I and II, (Wiley, New York, 1968).

[8] G. Jona-Lasinio, *The Renormalization Group: A Probabilistic View*, Il Nuovo Cimento, B**26**, 99-119 (1975).

[9] R. W. Batterman, *Why Equilibrium Statistical Mechanics Works: Universality and the Renormalization Group*, Philosophy of Science, **65**, 183–208 (1998); see also *Reduction and Renormalization*, talk presented at The R. and S. Boote Conference in Reductionism and Anti-Reductionism in Physics, University of Pittsburgh (April, 2006), URL: `http://philsci-archive.pitt.edu/archive/00002852/01/red-renorm.pdf`.

[10] M. E. Fisher, *The Renormalization Group in the Theory of Critical Behavior*, Reviews of Modern Physics **46**, 597–616 (1974).

[11] M. E. Fisher, "Scaling, Universality and Renormalization Group Theory", in *Critical Phenomena: Proceedings of the Summer School Held at the University of Stellenbosch, South Africa, January 18-29, 1982*, F.J.W. Hahne (Ed.) Springer Lecture Notes in Physics, Vol. 186, (Springer, Berlin, 1983), pp 1–139.

[12] E. R. Kandel, J.H. Schwarz and T.M. Jessel, *Principles of Neural Science*, 4th Edition, (Edward Arnold, London 2000).

[13] K.R. Popper and J. C. Eccles, *The Self and its Brain – An Argument for Interactionism*, (Springer, Heidelberg, Berlin, 1977).

[14] Ya. G. Sinai, *Probability Theory: An Introductory Course*, (Springer, Berlin, 1992)