

On the relevance of the maximum entropy principle in non-equilibrium statistical mechanics

Gennaro Auletta^{1,a}, Lamberto Rondoni^{2,3,4,b}, and Angelo Vulpiani^{4,5,6,c}

¹ University of Cassino, 03043 Cassino, Italy

² Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

³ INFN, Sezione di Torino, Via Pietro Giuria 1, 10125 Torino, Italy

⁴ Kavli Institute for Theoretical Physics China, CAS, Beijing 100190, P.R. China

⁵ Università di Roma “Sapienza”, Dipartimento di Fisica, Rome, Italy

⁶ CNR, Istituto Sistemi Complessi, P.le A. Moro 2, 00185 Rome, Italy

Received 1 March 2017

Published online 19 April 2017

Abstract. At first glance, the maximum entropy principle (MEP) apparently allows us to derive, or justify in a simple way, fundamental results of equilibrium statistical mechanics. Because of this, a school of thought considers the MEP as a powerful and elegant way to make predictions in physics and other disciplines, rather than a useful technical tool like others in statistical physics. From this point of view the MEP appears as an alternative and more general predictive method than the traditional ones of statistical physics. Actually, careful inspection shows that such a success is due to a series of fortunate facts that characterize the physics of equilibrium systems, but which are absent in situations not described by Hamiltonian dynamics, or generically in nonequilibrium phenomena. Here we discuss several important examples in non equilibrium statistical mechanics, in which the MEP leads to incorrect predictions, proving that it does not have a predictive nature. We conclude that, in these paradigmatic examples, an approach that uses a detailed analysis of the relevant aspects of the dynamics cannot be avoided.

1 Introduction

Statistical mechanics has been constructed, in the second half of the 19-th century, by Maxwell, Boltzmann and Gibbs on the basis of the assumed microscopic dynamics, and additional hypothesis (as ergodicity). On the other hand there is today a radically anti-dynamical point of view, according to which statistical mechanics were nothing else but a form of statistical inference, rather than a theory of objective physical

^a e-mail: jrduque@ucm.es

^b e-mail: dgomez@ucm.es

^c e-mail: carlos.mejia@upm.es

reality. Here with *statistical inference* it is understood the process of deducing properties of an underlying probability by means of some general criterion. Under this light, probabilities measure the degree of truth of a logical proposition about the state of the system, rather than describing the state of a system as such.

In this context, Jaynes [1–3] proposed the maximum entropy principle (MEP) as the general rule for finding the probability of a given event when only partial information is available. Let us briefly summarize the main points. If the mean values of m independent functions $f_i(\mathbf{x})$, where \mathbf{x} is a vector which describes the state of the system, are given:

$$c_i = \langle f_i \rangle = \int f_i(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \quad , i = 1, \dots, m, \quad (1)$$

the MEP rule determines the probability density ρ of the events compatible with these mean values, by maximising the “entropy”

$$H = - \int \rho(\mathbf{x}) \ln \rho(\mathbf{x})d\mathbf{x}, \quad (2)$$

under the constraints $c_i = \langle f_i \rangle$. For independent functions f_1, \dots, f_m we mean that it is not possible to find $a_1, \dots, a_m \neq 0$ such that

$$\sum_{j=1}^m a_j f_j(\mathbf{x}) = 0. \quad (3)$$

Using the maximization method of the Lagrange multipliers one easily obtains

$$\rho(\mathbf{x}) = \frac{1}{Z} \exp \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) \quad (4)$$

where $\lambda_1, \lambda_2, \dots, \lambda_m$ depend on c_1, c_2, \dots, c_m . For instance, for systems with a fixed number of particles subjected to the unique constraint that their mean energy is fixed, the MEP leads to the canonical distribution in a very simple fashion.

As a technical but rather important detail, we note that the above result holds only if \mathbf{x} is the vector of the canonical coordinates (i.e., positions and momenta of the particles). Analogously, for systems of varying numbers of particles, the grand canonical distribution is obtained by additionally constraining the mean number of particles.

Many find in these facts an unquestionable evidence for the validity of the MEP. We do not share such an opinion: the success of the MEP in deriving the correct probability distribution in equilibrium statistical mechanics is just a matter of fortunate coincidence, related to the choice of canonical coordinates. Thus, the weakest technical aspect of the MEP approach is the dependence of the results on the choice of the variables.

There is another and more important objection to the MEP: our ignorance cannot be credited to add knowledge about real phenomena. As a matter of fact, in spite of the optimistic claims of the MEP enthusiasts, to the best of our knowledge MEP has only produced different, although sometimes more elegant, derivations of previously known results. In other words, the MEP is analogous to other tools of statistical physics, that allow efficient calculations or summaries of known facts, but cannot be considered general inference principles.

For simplicity’s sake, consider a scalar random variable X , ranging over a continuum, whose probability distribution function is p_X . It is easy to realise that

the “entropy”

$$H_X = - \int p_X(x) \ln p_X(x) dx \quad (5)$$

is not an intrinsic quantity of the phenomena concerning X . With a different parametrization, i.e., using the coordinates $y = f(x)$ with an invertible function f , rather than x , the entropy of the same phenomenon would now be given by

$$H_Y = - \int p_Y(y) \ln p_Y(y) dy, \quad (6)$$

with $p_Y(y) = p_X(f^{-1}(y))/|f'(x = f^{-1}(y))|$. Therefore, one has

$$H_Y = H_X + \int p_X(x) \ln |f'(x)| dx, \quad (7)$$

so the MEP gives different solutions if different variables are adopted to describe the very same phenomenon.

In order to avoid such an unpleasant dependence on the choice of variables, Jaynes later proposed a more sophisticated version of the MEP, in terms of the relative entropy:

$$\tilde{H} = - \int \rho(\mathbf{x}) \ln \left[\frac{\rho(\mathbf{x})}{q(\mathbf{x})} \right] d\mathbf{x}, \quad (8)$$

where q is a given probability density. Of course, \tilde{H} depends on q ; but, at variance with the entropy, it does not depend on the chosen variables. On the other hand, one must decide how to select q , and this issue is equivalent to the problem of choosing the “proper variables”. Therefore, even this more elaborate method is non-predictive, and we see no reason to pursue the MEP approach further in the field of statistical mechanics. For a detailed discussion on the MEP in equilibrium statistical mechanics see [4–9].

The aim of the present paper is to discuss some non standard topics of statistical physics, namely fluid mechanics and non equilibrium problems, showing how the statistical features are determined by precise dynamical behavior and cannot be predicted (even at a qualitative level) by inference ideas as in the MEP.

In Section 2 the reader can find a short review of some delicate aspects of the entropy in statistical mechanics and its link with different interpretations of the probability.

In Section 3 we discuss the statistical mechanics of fluids: only in the inviscid case, which is for many aspects similar to the Hamiltonian systems, MEP is able to give the correct result. On the contrary, in the more interesting situation, of turbulent flows MEP is not able to select the correct statistical features which are selected only by some physical aspects of the dynamics.

Section 4 is devoted to the non equilibrium statistical mechanics. In some cases the MEP can predict the proper results but only using the relevant variables and constraints. Such assumptions, in terms of inference, are not natural at all. In particular the claimed success of the MEP for the fluctuation relations is due to serious confusion between formulae that are only apparently similar while, in reality, they describe completely different physical situations. The last section is devoted to our concluding remarks.

2 Entropy in statistical mechanics and interpretations of probability

Entropy has been originally introduced by Clausius in thermodynamics, then it has been extended in statistical mechanics, and interpreted as an indicator of “disorder”,

by Boltzmann and Gibbs. Shannon generalized the concept in information theory, after his seminal work entropy has been used (and abused) in many different contexts from biology to linguistic. In the following this section we briefly discuss some aspects of the entropy which, although simple, from a technical point of view, are rather subtle and can induce confusion.

Often, entropy is considered a thermodynamic quantity and therefore is supposed to have a kind of “surface” status relative to statistical-mechanical parameters. This view can appear bizarre when considering that the accepted formulation of entropy comes from the father of statistical mechanics, L. Boltzmann. It is therefore interesting to have a look at what happens in conditions that are not trivially ruled by thermodynamics, for instance at absolute zero temperature. As interesting example of such a situation we can mention that one studied in [10, pp. 609–610] relative to the probability distributions for an ice molecule.

It turns out that, even in absence of thermal motion and of any dissipative process, the hydrogen atoms can occupy six different locations, showing that entropy is not zero! If we insist on the thermodynamic character of entropy, the conclusion would be that Shannon’s entropy is capable to describe such a situation while Boltzmann’s one were not. However, Boltzmann’s entropy has a pure statistical–mechanical character and essentially deals with possible configurations and combinatorics of such configurations. And this evidently describes the situation that we are considering.

Moreover probability in statistical mechanics has again a fundamental, objective, and irreducible role. Although probabilities need a consistent theoretical framework in order to be computed at all, a proper theory has precisely the function to help us to compute the useful probabilities, that is those that are nearer to the statistics that are in fact observed (without necessarily resorting to a frequentist view of probability). So, the theoretical character of probabilities does not deprive them of objective character at all. Classical determinism, according to which any possible event is either certain or impossible is in fact falsified. This is an important commonality with quantum mechanics (what is also obvious for the atomic-molecular scale), and in fact recently the relevance of Boltzmann’s constant for quantum mechanics has been stressed [11, 12].

A popular way to formulate the second law of thermodynamics is to say that the entropy increases in time; such a statement deserves careful considerations, in particular how to associate an entropy-like quantity to the state of a physical system [13]. Consider a Hamiltonian system with N weakly interacting particles, its microscopic state, in the so called Γ space, is given by a vector $\mathbf{x} = (\mathbf{q}_1, \dots, \mathbf{q}_N, \mathbf{p}_1, \dots, \mathbf{p}_N)$, where \mathbf{q}_n and \mathbf{p}_n , are the position and the momentum of the n -th particle respectively. From the probability distribution in the Γ -space $\rho(\mathbf{x}, t)$ at time t , following Gibbs, we can introduce an entropy

$$S_G(t) = - \int \rho(\mathbf{x}, t) \ln \rho(\mathbf{x}, t) d\mathbf{x}.$$

Using the Liouville theorem it is easy to show that $S_G(t)$ must be constant.

In order to observe an increase over time for S_G like quantities, one can introduce a coarse-graining of the Γ -space, amounting to consider a partition of the phase space in cells of size ϵ . In this way one obtains a coarse-grained Gibbs entropy, which turns to be an increasing function of time, such a behaviour is surely interesting in a dynamical system context. On the other hand numerical results show that even such a coarse-grained Gibbs entropy remains constant up to a crossover time $t_*(\epsilon) \sim \ln(1/\epsilon)$, after which it starts increasing. This ϵ -dependence indicates that this growth has no thermodynamic meaning and it is originated by the coarse-graining procedure [14].

Another way to introduce the concept of entropy in weakly interacting systems is due to Boltzmann. Instead of the probability distribution in the Γ -space, let us

consider the one particle distribution function in the μ space:

$$f(\mathbf{r}, t) = \frac{1}{N} \sum_{n=1}^N \delta(\mathbf{r} - \mathbf{r}_n(t))$$

where $\mathbf{r}_n = (\mathbf{q}_n, \mathbf{p}_n)$. This distribution, under proper conditions that include the large system, $N \rightarrow \infty$, limit (not required by probability densities in phase space), can be replaced at a mesoscopic level by a continuous distribution [15]. Using the same symbol f for this continuous distribution, we can introduce the Boltzmann entropy

$$S_B(t) = - \int f(\mathbf{r}, t) \ln f(\mathbf{r}, t) d\mathbf{r}.$$

It is easy to realise that at equilibrium the two entropies are equivalent: $S_G \simeq NS_B$. On the other hand from the celebrated H theorem we know that, under suitable conditions, S_B increases in time [16].

At a first glance the two definitions of entropy sound rather similar, on the other hand their dynamical behaviours are rather different. Let us comment about the fact that the two definitions reflect some conceptual differences how to consider probability. It is well known that there are several competing interpretations of the actual meaning of the probability. Here our interest is limited to the link between probability and real world, in particular our interest is to investigate the following topic: what is the physical link between the probabilistic computations (i.e., the averages over an ensemble) and the actual results obtained in laboratory experiments which, a fortiori, are conducted on a single realization (or sample) of the system under investigation?

Roughly speaking the different points of view can be classified in two large classes: the subjective interpretation and the objective one. According to the subjective interpretation probability is a degree of belief; one of the most influential supporter of such a point of view is Jaynes with his idea that the behaviour of physical systems is governed by the degree of belief of the observer. In the objective interpretation the probability of an event is determined by the physics of the systems and not the lack of information of the observer [17].

Let us go back to the entropy S_G : one needs $\rho(\mathbf{x}, t)$, in other words an ensemble. It is easy to realize that $\rho(\mathbf{x}, t)$, and therefore $S_G(t)$, is accessible only in numerical experiments with systems composed by few degrees of freedom. On the contrary, if $N \gg 1$, the one particle distribution function $f(\mathbf{r}, t)$, and therefore $S_B(t)$, can be seen as an empirical distribution, is a well defined macroscopic observable and can be, in principle, measured in a single system; actually this is done in numerical simulations [13, 16].

3 Statistical fluid mechanics and maximum entropy principle

In spite of the fact that, in general, a fluid (even in absence of viscosity) does not obey Hamiltonian equations, it is easy to develop an equilibrium statistical mechanical treatment for the Euler equation in perfect analogy with the micro-canonical formalism used in standard Hamiltonian systems.

Let us consider a perfect fluid, i.e., with zero viscosity, and without external forcing, in a cube of edge L with periodic boundary conditions, so that the velocity field can be expanded in Fourier series as

$$u_j(\mathbf{x}, t) = L^{-3/2} \sum_{n_1, n_2, n_3} e^{i(k_1 x + k_2 y + k_3 z)} v_j(\mathbf{k}, t) \quad (9)$$

where

$$\mathbf{k} = \frac{2\pi}{L}(n_1, n_2, n_3)$$

with n_j integer numbers. We introduce an ultraviolet truncation $v_j(\mathbf{k}) = 0$ for $|\mathbf{k}| > K_M$, being K_M the maximum allowed wave vector.

Because of the incompressibility condition ($\nabla \cdot \mathbf{u} = 0$) and the fact that the velocity field $\mathbf{u}(\mathbf{x}, t)$ is real the variables $\{v_j(\mathbf{k}, t)\}$ are not independent, e.g., one has

$$\sum_{j=1}^3 k_j v_j(\mathbf{k}, t) = 0 \text{ and } v_j(-\mathbf{k}, t) = [v_j(\mathbf{k}, t)]^*, \quad (10)$$

where \star denotes complex conjugation. Therefore it is useful to introduce a new set of variable $\{X_n(t)\}$ replacing $v_j(\mathbf{k}, t)$ and obeying an ordinary differential equation:

$$\frac{dX_n}{dt} = \sum_{m,\ell} M_{n,m,\ell} X_m X_\ell, \quad n = 1, 2, \dots, N \sim K_M^3. \quad (11)$$

From the Euler's equation we have the following properties: where $M_{n,m,\ell} = M_{n,\ell,m}$ and $M_{n,m,\ell} + M_{m,\ell,n} + M_{\ell,n,m} = 0$: for details see [18, 19]. Thanks to the introduction of the ultraviolet truncation, we have a finite system of equations that avoid the infinite energy problems of the classical field theory.

Since equation (11) conserves the volume in the phase space (Liouville theorem)

$$\sum_n \frac{\partial}{\partial X_n} \frac{dX_n}{dt} = 0 \quad (12)$$

and in addition one has the (energy) conservation law

$$\frac{1}{2} \sum X_n^2 = E,$$

it is straightforward, following the usual approach of equilibrium statistical mechanics, to derive the microcanonical distribution:

$$P_{mc}(\{X_n\}) \propto \delta\left(\frac{1}{2} \sum X_n^2 - E\right). \quad (13)$$

In addition, the $N \rightarrow \infty$ limit yields the canonical distribution

$$P_c(\{X_n\}) \propto \exp\left[-\left(\frac{\beta}{2} \sum X_n^2\right)\right] \quad (14)$$

and therefore

$$\langle X_n^2 \rangle = \frac{2E}{N} = \frac{1}{\beta}. \quad (15)$$

The previous procedure can be easily generalized to the two-dimensional case in which there is a second conserved quantity, the enstrophy (the quantity related to the kinetic energy in the flow model that corresponds to dissipation effects in the fluid):

$$\Omega = \frac{1}{2} \sum_n k_n^2 X_n^2. \quad (16)$$

Because of this, the microcanonical distribution should be defined on the surface in which both energy and enstrophy are constant, and in the large N limit, we have the canonical distribution

$$P_c(\{X_n\}) \propto \exp \left[- \left(\frac{\beta_1}{2} \sum X_n^2 + \frac{\beta_2}{2} \sum k_n^2 X_n^2 \right) \right] \quad (17)$$

and therefore

$$\langle X_n^2 \rangle = \frac{1}{\beta_1 + \beta_2 k_n^2} . \quad (18)$$

Detailed numerical simulations show that systems described by inviscid ordinary differential equations, such as equation (11), with quadratic invariants, for which the Liouville theorem holds, are ergodic and mixing if N is large. Then arbitrary initial distributions of $\{X_n\}$ evolve towards the Gaussian (14) or (17), see [19].

In the inviscid case, the Liouville theorem implies that the “natural” variables are $\{X_n\}$, and the success of the MEP to derive the canonical distribution is quite obvious. The reason is the same as for the statistical mechanics of Hamiltonian systems. We additionally remark that in usual statistical mechanics, one only uses the energy conservation and the Liouville theorems, while the full Hamiltonian structure plays no role.

Let us now discuss the more interesting case of real fluids, where a viscosity and forcing are present. Particularly interesting is the fully developed turbulence, where the Reynolds number $R_e = UL/\nu$ (being U and L the typical velocity and length respectively) is very high. At variance with naive expectations on the statistical features of turbulence, based on the incorrect assumption that $R_e \rightarrow \infty$ is equivalent to $\nu = 0$, the scenario is very different [18]. In 3D, instead of equipartition, we have Kolmogorov’s law i.e. $E(k_n) \sim k_n^2 \langle X_n^2 \rangle \sim k_n^{-5/3}$, therefore

$$\langle X_n^2 \rangle \sim k_n^{-11/3} . \quad (19)$$

The previous law can be understood in terms of a cascade mechanism: see [18,19].

Note that what may seem a natural constraint for the application of the MEP

$$\sum_n \langle X_n^2 \rangle = \text{const}, \quad (20)$$

leads to equation (15) i.e., to the same result of the inviscid case. Let us stress that it is neither simple nor natural to introduce other constraints. For instance, the dissipation due to the viscosity and to the forcing terms, makes the Liouville theorem fail and leads to fractal phase space distributions for which no “natural” variables exist.

3.1 Statistical features of turbulent models

In order to understand of the difference between the cascade mechanism and the “equipartition” scenario a numerical study is unavoidable. On the other hand, a numerical simulation of the Navier–Stokes equations in the limit $R_e \gg 1$ is a prohibitive task. If the interest is only for a study of the scaling behavior, one can use simplified dynamical models, the so called shell models (SM), which, in spite of their apparent simplicity, reproduce many statistical features observed in experiments and in detailed numerical simulations [19–21].

The basic idea of the SM is to implement a dynamical (energy or other quantities) cascade model in terms of a set of complex variables u_n , $n = 1, \dots, N$ representing the

velocity fluctuations in a shell of wave-numbers $k_n < |\mathbf{k}| < k_{n+1}$. The wave-numbers are chosen geometrically spaced $k_n = k_0 2^n$ therefore the number of variables needed to describe the inertial range physics, is not too large. In this way, the spatial and vectorial structure of the original problem is completely disregarded. Then, some insights are used to derive the equations ruling the set of variables $\{u_n\}$. A basic source of inspiration is the Navier-Stokes equation (NSE) written in Fourier space, where the modes interact in triads (see e.g., Eq. (11)): only three modes are involved at the same time. In this way, we simplify the complexity of the equations by retaining the triad structure and eliminating some interactions. Due to the hierarchical organization of the characteristic times associated with the different scales, we can assume that only close modes, i.e., variables referring to close scales, can interact. The justification for this is that distant modes (say k_n and k_m with $|m - n| \gg 1$) have so different timescales that the resulting interaction would be very weak. This assumption is known as the hypothesis of locality of the cascade, and can be substantiated with refined analysis of the NSE.

According to the previous ideas we can introduce a set of ordinary differential equations:

$$\frac{du_n}{dt} = -\nu k_n^2 u_n + g_n^{(\alpha)}(u_n, u_{n\pm 1}, u_{n\pm 2}) + f_n, \quad (21)$$

where f_n is the external forcing, the term $-\nu k_n^2 u_n$ corresponds to the dissipation, while the term $g_n^{(\alpha)}(\dots)$ includes the nonlinear terms, and the parameter α determines the conservation laws in the inviscid limit $\nu = 0$, $f_n = 0$:

$$\frac{du_n}{dt} = -\nu k_n^2 u_n + f_n + ik_n \left(a_n u_{n+1}^* u_{n+2}^* + \frac{b_n}{2} u_{n-1}^* u_{n+1}^* + \frac{c_n}{4} u_{n-1}^* u_{n-2}^* \right), \quad (22)$$

with $n = 1, \dots, N$, $b_1 = b_N = c_1 = c_2 = a_N = a_{N-1} = 0$.

Given the conservation of energy $\sum_n |u_n|^2$ when $\nu = f_n = 0$, one has the constraint $a_n + b_{n+1} + c_{n+2} = 0$, and the time scale can be fixed applying the condition $a_n = 1$. This leaves one free parameter δ :

$$a_n = 1, \quad b_n = -\delta, \quad c_n = -(1 - \delta). \quad (23)$$

In the inviscid limit, the systems possesses a second conserved quantity:

$$\sum_n k_n^\alpha |u_n|^2, \quad (24)$$

where α and δ are linked by the relation $2^\alpha = 1/(1 - \delta)$ [19]. The cases $\delta = 1/2$ and $\delta = 5/4$ correspond the $3d$ and $2d$ turbulence respectively.

In spite of their (apparent) naive character, the shell models are non trivial at all and maintain all the difficulties of the NSE. Remarkably for $\delta = 1/2$ the shell model shows the same rich statistical features observed in labs and direct numerical simulations of the NSE, e.g., the anomalous scaling of the structures functions and the shapes of the probability distribution of many relevant quantities. Let us note that the agreement holds also at the quantitative level.

The great advantage of shell models is that the number of shells N necessary to mimic the cascade mechanism of fully developed turbulence is relatively small, because of the geometrical progression in k_n we roughly have $N \sim \ln(R_e)$. We have thus a chaotic dynamical system with a reasonably small number of degrees of freedom where methods of deterministic chaos can be used to link the statistical description to the dynamical properties.

In the past years shell models attracted the attention of many scientists with different aims: the possibility to perform detailed numerical computation on a model

for the energy cascade to test ideas or conjectures, e.g., in the context of predictability. Also, they have been used to investigate analytic methods (to test some ideas for the closure problem), developing rigorous results, understanding the link between dynamical properties in phase space and more standard quantities (in traditional turbulent literature) such as structure functions and velocity probability distribution.

Although only the cases $\delta = 1/2$ and $\delta = 5/4$ correspond to real physical situations ($3D$ and $2D$, respectively), it is interesting to study the model also for other values of δ : see [19]. Let us discuss only the case $\delta > 1$ corresponding to real α .

In order to determine the main statistical features, we can follow two different statistical arguments:

- a) equipartition, i.e., $(k_n^\alpha + \text{const.})\langle |u_n|^2 \rangle = \text{const}$ which, for large k_n implies $\langle |u_n|^2 \rangle \sim k_n^{-2\zeta}$ where $\zeta = \alpha/2$;
- b) cascade á la Kolmogorov (see e.g., [19]), obtaining $\langle |u_n|^2 \rangle \sim k_n^{-2\zeta}$, where $\zeta = (\alpha + 1)/3$.

Actually, numerical simulations (see [19,20]) show, apart small corrections due to intermittency, $\zeta = (\alpha + 1)/3$ if $\alpha < 2$, and $\zeta = \alpha/2$ for $\alpha > 2$, i.e.,

$$\zeta = \max\left\{\frac{\alpha + 1}{3}, \frac{\alpha}{2}\right\}. \quad (25)$$

In short, we can say that for $\alpha < 2$ the most important mechanism is the cascade (à la Kolmogorov), while for $\alpha > 2$ the equipartition mechanism has the leading role for the scaling. Let us present the physical argument [19]: neglecting intermittency we assume the simple scaling $u_n \sim k_n^{-h}$, therefore by dimensional arguments, the typical time at scale k_n is $\tau_n \sim u_n k_n \sim k_n^{h-1}$. Assuming a generalized α -enstrophy cascade one has a constant rate for the generalized α -enstrophy transfer, i.e., $k_n^\alpha |u_n|^2 / \tau_n \sim \text{const}$. Therefore one obtains $h = (1 + \alpha)/3$ corresponding to $\zeta = (\alpha + 1)/3$.

In the previous argument one has

$$\tau_n \sim k_n^{(\alpha-2)/3},$$

such a scaling is contradictory for $\alpha \geq 2$, because it implies that the turn-over time τ_n does not decrease as k_n^{-1} decreases. Therefore we have an unrealistic result: it is not possible to stop the cascade with a dissipative mechanism whose characteristic time is $\tau_n^{(d)} \sim k_n^{-2}$, i.e., at variance with the case $\alpha < 2$, it is not possible to find a $k_{diss} = k_{n^*}$ such that

$$\tau_{n^*} \sim \tau_{n^*}^{(d)}.$$

As consequence of the failure of the cascade mechanism, for $\alpha \geq 2$, it is reasonable to expect an equilibrium statistical scenario, with $\zeta = \alpha/2$.

The previous phenomenological argument is well confirmed by numerical computations, see [19,20].

The above results on the dependence on ζ from α originate from genuine physical arguments and it is hard to believe that they could be obtained by mere inference arguments.

4 Non-equilibrium examples

In this section we discuss the MEP approach in the case of non-equilibrium systems. Dewar [22,23] claimed that the MEP can be used to obtain the probability distributions for general non-equilibrium systems.

For sake of simplicity, let us consider discrete times, 1,2,3,... and let

$$\mathcal{T}_N = \{x_1, x_2, \dots, x_N\} \quad (26)$$

be a trajectory segment of length N in the phase space or state space of a given system of interest. One would like to identify the steady state probability density $p(\mathcal{T}_N)$ about this trajectory segment in the state space of trajectory segments. The method stemming from the MEP relies on the maximization of the corresponding “entropy”:

$$H_N = - \int p(\{x_1, \dots, x_N\}) \ln p(\{x_1, \dots, x_N\}) dx_1 \dots dx_N \quad (27)$$

under the constraints concerning M observables:

$$c_{j,N} = \langle f_{j,N} \rangle, \quad j = 1, 2, \dots, M \quad (28)$$

where $f_{j,N} = f_{j,N}(\mathcal{T}_N)$: the scalars $c_{j,N}, f_{j,N}$ play the same roles as c_j, f_j did in Section 1, and the notation stresses the fact that they refer to trajectories of length N in the original phase space or state space. The MEP immediately leads to

$$p(\mathcal{T}_N) = e^{-\sum_{j=1}^M \lambda_j f_{j,N}(\mathcal{T}_N)} \quad (29)$$

where the values $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$ are determined by those of $\{c_{1,N}, c_{2,N}, \dots, c_{M,N}\}$ in equation (28).

In this respect, the MEP method does not differentiate equilibrium from non-equilibrium steady states; once the trajectory of interest is identified, everything proceeds in the same way for both situations. This is indeed in line with considering the maximum entropy as a inference technique, which is then expected to work regardless of the physics and of the specific properties of the objects under investigation. The difficulties one meets in such a completely general approach to non-equilibrium systems are as above:

- a) the identification of the stationary state, i.e., of the suitable variables for describing it;
- b) the identification of the observables, i.e., of the relevant functions $\{f_j\}$ in equation (28).

Cover and Thomas [24] in their well known book, express these ideas quite plainly:

Implicit in the use of the maximum entropy methods in physics is a sort of AEP (asymptotic equipartition property) that says that all the micro states are equally probable.

In Section 12.6 of [24], the reader can find a clear discussion of Burg’s maximum entropy theorem, which states that the stochastic process $\{x_1, x_2, \dots, x_N\}$ satisfying

$$\langle x_n x_{n+k} \rangle = C_k, \quad k = 1, 2, \dots, p, \quad (30)$$

where the correlation function C_k is known, and enjoying the maximum entropy rate

$$h = \lim_{N \rightarrow \infty} \frac{1}{N} H_N, \quad (31)$$

is the Gaussian Markov process obeying

$$x_n = \sum_{k=1}^p A_k x_{n-k} + \sigma z_n, \quad (32)$$

with $\{z_n\}$ i.i.d. Gaussian variables of zero mean and unitary variance, and $\{A_k\}$ and σ chosen so that equation (30) is satisfied. Then, one has

$$p(\{x_1, x_2, \dots, x_N\}) = \frac{1}{K} \exp -\frac{1}{2\sigma^2} \sum_n \left(x_n - \sum_p A_p x_{n-p} \right)^2, \quad (33)$$

where K is the normalization constant.

Arguably, in the simplest case one assumes

$$\langle x_n x_{n+k} \rangle = a^k \langle x^2 \rangle, \quad \text{with } 0 < a < 1, \quad (34)$$

and the process maximizing h is a discrete time Langevin equation, of form:

$$x_n = ax_{n-1} + \sigma z_n \quad (35)$$

with $\sigma^2 = \langle x^2 \rangle (1 - a^2)$. As well known, the corresponding probability density p is then given by [29]

$$p(\{x_1, \dots, x_N\}) = \frac{\exp \left[-\frac{1}{2\sigma^2} \sum_n (x_n - ax_{n-1})^2 \right]}{K}. \quad (36)$$

Despite being technically simple, this result reveals the serious limitations in which the MEP inevitably incurs.

For instance, Dewar [22, 23] using the MEP approach obtains

$$p(\{x_1, x_2, \dots, x_N\}) = e^{-\sum_j \sum_n \lambda_j g_j(x_n)} \quad (37)$$

but with values $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$ derived from constraints that are sums of functions of the variable x at a given time. Consequently, one cannot account for a sum of functions of x_n and x_{n-1} . One may insist, and define the state at time n in terms of a two components array such as $\mathbf{y}_n = (x_n, x_{n-1})$. In that case, indeed, the MEP approach might work, if proper constraints are imposed, and one could claim that this is done because the proper variables have been used.

On the other hand, the state at time n for the present system is merely given by x_n , and there is no *a priori* reason to adopt $\mathbf{y}_n = (x_n, x_{n-1})$ as a description of the state. One could equally reasonably opt for any other pair of variables. In a sense, such choices would be like deciding that the harmonic oscillator is described by $x, dx/dt$ and d^2x/dt^2 , instead of $x, dx/dt$ only.

Let us recall in such a context the celebrated Onsager-Machlup work [25] on fluctuations and irreversible processes, where the authors pose a difficult question:

how do you know you have taken enough variables, for it to be Markovian?

Similarly, Ma [26], page 29, observes:

the hidden worry of thermodynamics is: we do not know how many coordinates or forces are necessary to completely specify an equilibrium state.

4.1 Fluctuation relations

Dewar [22, 23] used the information theoretic approach to non-equilibrium statistical mechanics also in order to derive one of the most popular results of the past decades: the fluctuation relations (FR) which deals with the probabilities of a trajectory and its (time) reversed one. This is a symmetry relation of the probability of second law

“violating” phase-space paths [27–29]. Dewar’s derivation seems to imply that the FR is a generic property of the MEP’s probability distributions, involving constraints on anti-symmetric functions, independently of any physical interpretation that may be associated to the phenomenon and to the constraints. Physically, it would then suffice to apply the MEP to the entropy production of those macroscopic fluxes that vary under the imposed constraints, and that would amount to selecting the most probable macroscopic flux configuration. In this case, one denotes by f_k the thermodynamic fluxes contributing to the entropy production.

Let $\mathcal{T}_N^{(+)}$ be a trajectory of length N and probability $p_{\mathcal{T}_N^{(+)}}$, along which there is a positive entropy production, expressed by $\sigma = \sum_{k=1}^m \lambda_k f_k(\mathcal{T}_N^{(+)})$. Such a trajectory can be paired with a trajectory $\mathcal{T}_N^{(-)}$ of probability $p_{\mathcal{T}_N^{(-)}}$, corresponding to the opposite entropy production, $-\sigma$. Using MEP, Dewar then obtains the following relation — cf. equation (8) in reference [22, 23] —

$$\frac{p_{\mathcal{T}_N^{(+)}}}{p_{\mathcal{T}_N^{(-)}}} = \exp \left\{ \sum_{k=1}^m \lambda_k f_k(\mathcal{T}_N^{(+)}) \right\}, \quad (38)$$

which is apparently very general, because rather than entropy production for a non-equilibrium thermodynamic systems, one could have considered any process with n outcomes $\{1, \dots, n\}$, whose events can be grouped in pairs $(i+, i-)$, such that the f_k ’s obey $f_k(i-) = -f_k(i+)$. The result would have been identical, proving the incredibly general applicability of the FR.

Dewar then observes that: *a common explanation for these relationships lies in the hypothesis that the trajectories have a Gibbs-type probability distribution. Maximal Entropy provides the natural formalism in which Gibbs-type distributions emerge, whether or not they refer to physical systems. Thus the fluctuation theorem is not confined to physical systems alone but arises in a (potentially large) class of statistical inference problems involving constraint functions which are anti-symmetric.*

To connect Dewar’s result with the FR of non-equilibrium statistical physics, one has to take a sample space whose n elements are the possible trajectories of a non-equilibrium system. This implies that the state space (for a stochastic process) is finite, or that the phase space (for a deterministic system) admits a finite generating partition.

Granting all that, it nevertheless appears that Dewar’s equation (38) does not distinguish the numerous different situations that may arise, and always yields the same expression (38) even in the cases in which it is incorrect. Therefore, this MEP approach has no predictive value. In particular, equation (38) incurs in a systematic error which is far from harmless, because it puts on the same footings two physically completely different questions and the correspondingly different experiments:

- 1) Measurements concerning the properties of non-equilibrium steady states;
- 2) Measurements concerning the properties of equilibrium states.

This formally appears in the fact that, *even when it does hold*, a steady state FR does not look like equation (38) in general, but it contains a correction term c_N that must turn negligible when N grows:

$$\frac{p_{\mathcal{T}_N^{(+)}}}{p_{\mathcal{T}_N^{(-)}}} = \exp \left\{ \sum_{k=1}^m \lambda_k f_k(\mathcal{T}_N^{(+)}) + d_N \right\}. \quad (39)$$

In the standard cases in which the steady state FR holds, d_N is of order $O(1)$ compared to the order $O(N)$ of the sum in equation (39), hence it is indeed negligible for large

N . However, d_N is related to the decay of correlations of microscopic events in the steady state, and when the relevant correlations do not decay sufficiently fast, it may get large with N producing expressions that do not resemble equation (38) — cf. [28–32] — at variance with the MEP approach.¹

As a matter of fact, a relation like equation (38) lacking the correction term d_N , can be verified, but as a transient relation. Transient relations are very interesting tools, used to obtain equilibrium properties of given collections of system, by doing non-equilibrium work on them [29]. In a sense, transient FR close the circle with the Fluctuation Dissipation Theorem which does the opposite, obtaining non-equilibrium properties from equilibrium experiments.

One could then argue that a correction of order $O(1)$ on a term of order $O(N)$ should be neglected in general, and that equation (38) could be accepted *in practice* in all circumstances. However, this is a gross error. Apart from the above observations, there are at least three further major differences in the physics described by steady state and transient FR:

- there is no indication in the MEP procedure that N should be large, therefore the accuracy of the supposed approximation of the correct FR cannot be estimated;
- the probability p appearing in the transient FR is the equilibrium probability and not the steady state probability, so that p in equation (38) intended as a transient relation and p in equation (39) are totally different objects;
- transient FR describe the statistics of different experiments starting in the same equilibrium state but with different initial microscopic state, as in the case of protein stretching or of colloidal particles dragged in water. Consequently, transient FR do not need to describe any given single object, in general, and they do not even need to tend to any steady state expression when N grows. Differently, steady state FR describe the fluctuations of a single non-equilibrium system in its steady state [27–29].

Furthermore, the most serious difficulty lies again with the choice of the functions f_k . Analogously to the previous general discussion, also in the case of the fluctuation relations one should know beforehand the correct variables by which the state of a system must be described, as well as the relevant observables. Unfortunately, in many circumstances a proper set of variables does not even exist [33], and one may pass from a situation in which the steady FR holds (correlations decay fast and d_N turns negligible with increasing N) to one in which it does not (d_N remains comparable to the other terms) by merely changing parameters which play no role in the MEP approach [34,35].

4.2 A working example

The above remarks for deterministic dynamics have stochastic counterparts. Therefore, let us conclude this section considering a simple model, in which similar difficulties are encountered, as first pointed out by Farago for systems in unbounded potentials [36]. In particular, let us consider an overdamped Langevin process,

¹ Note that the N is not required to be large for the system to reach a steady state; d_N is present within the steady state dynamics. Also, one cannot consider the collection of infinitely long trajectories ($N = \infty$) because, apart from those corresponding to the average entropy production, their probabilities vanish and the ratio on the left hand side of (39) has no meaning

describing a Brownian particle, dragged in a liquid by a moving harmonic potential with a constant velocity v^* , which is relevant e.g., for the optical trap experiment [37]:

$$\frac{dx(t)}{dt} = -[x(t) - x^*(t)] + \zeta(t) . \quad (40)$$

Here $x(t)$ is the position of the particle at time t , $x^*(t) = v^*t$ the position of the minimum of the potential, $\zeta(t)$ is a white noise term representing the thermal bath, and $k_B T = 1$. Then, the work done in a time τ is

$$W_\tau = -v^* \int_0^\tau [x(t) - x^*(t)] dt . \quad (41)$$

In this context, Van Zon and Cohen [38] considered the energy balance

$$W_\tau = Q_\tau + \Delta U_\tau \quad (42)$$

where Q_τ is the dissipated heat and ΔU_τ the potential energy of a colloidal particle. They then observed that in a comoving frame W_τ is Gaussian with variance $2\langle W_\tau \rangle$. As this property persists asymptotically in τ , they could conclude that the steady state FR holds for the total work.

Differently, the PDF of the potential energy is exponential at equilibrium, $P(\Delta U) \sim \exp(-\text{const} \cdot \Delta U)$, and is expected to remain exponential even away from equilibrium. Consequently, the small fluctuations of heat are expected to coincide with those of the total work, because the contribution of the potential energy is only $\mathcal{O}(1)$, while large heat fluctuations are more likely to be generated by large fluctuations of the potential energy, thus they are not distributed like work.

As a result, the expression in the large τ limit for the heat FR takes the standard form

$$\frac{P_\tau(Q_\tau)}{P_\tau(-Q_\tau)} \approx e^{Q_\tau} \quad (43)$$

only for $Q_\tau \in [0, \langle Q \rangle)$. For $Q_\tau \in [\langle Q \rangle, 3\langle Q \rangle)$, there is a complicated nonlinear function of Q_τ and W_τ in the exponential, and for $Q_\tau > 3\langle Q \rangle$ one eventually obtains

$$\frac{P_\tau(Q_\tau)}{P_\tau(-Q_\tau)} \approx e^{2\langle Q \rangle} \quad (44)$$

where $\langle \cdot \rangle$ is the steady state average.

This result, due to the insufficiently rapid decay of the PDF of heat, as opposed to that of the PDF of work, means that two perfectly analogous quantities from the MEP standpoint, the work and the dissipated energy, are in fact described by two substantially different FR, at variance with the MEP predictions.

Furthermore, Baiesi et al. [39] generalized the result of [38], providing necessary conditions for the potential V and for its motion $x^*(t)$, which are required by the total work to satisfy the steady state FR. In particular, numerical tests showed that the steady state FR does not hold for the total work if x^* moves at constant velocity and V is not symmetric. Similar observations are reported in [40, 41].

As one may obtain non-symmetric potentials by changing one parameter in the model of [38], without affecting the parity of the total work, here we have another example in which the MEP approach is bound to make incorrect predictions.

The fact is that there are infinitely many different forms for the FT, each of which depends on specific details of the systems under consideration. It would be quite a surprise that any method generically based on an equipartition property treats

properly such a plethora of different situations, especially considering that a typical feature of non-equilibrium systems, even close to equilibrium, is the absence of equipartition [42]. As shown in many papers, this fact affects in vastly different fashions different observables and, to date, we have no way to point out which observables are most affected, except by direct investigation.

5 Concluding remarks

Apparently, the MEP approach may look like an elegant and powerful way to make statistical inference in non-equilibrium situations. However, detailed analysis reveals a serious difficulty, which makes non-predictive the MEP method: it allows the derivations of the correct expressions only when they are already known.

However, it is possible to see that entropy is a rather subtle concept which cannot be used without a critical approach. In the non-equilibrium cases, one cannot proceed without knowledge of the dynamics, which tends to be highly complex. In other words, the constraints one should impose, even in the cases in which this can be done, should be related to the dynamics.

It is worth mentioning that the MEP approach is also adopted in situations more complex than those concerning physical problems such as the ones discussed above. Biology, for instance, provides countless examples in which the details of the relevant dynamics are not understood, hence it is often claimed that uninformed inference, of the MEP kind, is necessary. However, the maximization of the entropy is not appropriate to describe a living organism, since living organisms are not in equilibrium with the environment and they are characterized by a degree of order higher than that of the environment [43, 44]. This kind of order persists in time thanks to the energy and matter exchange of one organism with its environment. In non-equilibrium situations, correlations prevent the system from reaching the possible maximum entropy. Studies such as [46, 47] show that biomolecules are in general not in a state of maximal entropy precisely due to correlations among different components, as it is evident e.g., in the emergence of the so-called ternary or quaternary structure of proteins from the initial codified segments of amino acids, that ultimately allow biological functions.

One may thus argue that the MEP should be replaced by an analogous inference principle suitable to characterize this exchange. For instance, Prigogine's minimum entropy production principle correctly describes the system-environment exchange for stationary states in the linear regime of irreversible thermodynamics. The principle asserts that the steady-state configuration minimizes the entropy production [45]. The principle also suggests that the evolution will promote organisms that minimize their entropy production in their own environment, rather than maximizing the entropy. These organisms should turn out to be precisely those that are able to exert more control on the environment. However, the minimum entropy production principle does not possess an absolute generality. It fails as a system is taken farther and farther away from equilibrium, i.e., when it is driven towards higher and higher dissipations. Therefore it suffers from the same difficulties of the MEP and, analogously to the MEP, it cannot be used as a general inference principle.

References

1. E.T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**, 620 (1957)
2. E.T. Jaynes, Information theory and statistical mechanics II, *Phys. Rev.* **108**, 171 (1957)
3. E.T. Jaynes, Information Theory and Statistical Mechanics, in K. Ford (ed.), *Statistical Physics* (Benjamin, New York, 1963), pp. 181–218

4. S. Chibbaro, L. Rondoni, A. Vulpiani, *Reductionism, Emergence and Levels of Reality: The Importance of Being Borderline* (Springer, Heidelberg, 2014)
5. S. Chibbaro, L. Rondoni, A. Vulpiani, On the foundations of statistical mechanics: ergodicity, many degrees of freedom and inference, *Comm. Theor. Phys.* **62**, 469 (2014)
6. J. Uffink, Can the maximum entropy principle be explained as a consistency requirement?, *Stud. History Philos. Mod. Phys.* **26**, 223 (1995)
7. J. Uffink, The constraint rule of the maximum entropy principle, *Stud. History Philos. Mod. Phys.* **27**, 47 (1996)
8. A. Shimony, The Status of the Principle of Maximum Entropy, *Synthese* **63**, 35. *Search for a Naturalistic Point of View*, Cambridge, University Press.
9. K. Friedman, A. Shimny, Jaynes's maximum entropy prescription and probability theory, *J. Stat. Phys.* **3**, 381 (1971)
10. P.W. Atkins, J. De Paula, *Physical Chemistry* (Oxford, University Press, 2002, 2nd ed. 2006)
11. A.C. Elitzur, Locality and indeterminism preserve the second law, *Phys. Lett.* **A167**, 335 (1992)
12. A.C. Elitzur, S. Dolev, Quantum Phenomena Within a New Theory of Time, in Elitzur, A.C., Dolev, S., and Kolenda, N. (eds.), *Quo Vadis Quantum Mechanics?* (Springer, Berlin, 2005) pp. 325–49.
13. J.L. Lebowitz, Boltzmann's entropy and time's arrow, *Phys. Today* **46**, 32 (1993)
14. M. Falcioni, L. Palatella, S. Pigolotti, L. Rondoni, A. Vulpiani, Initial growth of Boltzmann entropy and chaos in a large assembly of weakly interacting systems, *Phys. A* **385**, 170 (2007)
15. O.E. Lanford, III, in: J. Moser (ed.), *Dynamical Systems, Theory and Applications*, Lecture Notes in Physics (Springer, Berlin, 1975), Vol. 38
16. L. Cerino, F. Cecconi, M. Cencini, A. Vulpiani, The role of the number of degrees of freedom and chaos in macroscopic irreversibility, *Physica A* **442**, 486 (2016)
17. A.L. Kuzemsky, Probability information and statistical physics, *Int. J. Theor. Phys.* **55**, 1378 (2015)
18. U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov* (Cambridge University Press, 1995)
19. T. Bohr, M.H. Jensen, G. Paladin, A. Vulpiani, *Dynamical Systems Approach to Turbulence* (Cambridge University Press, 1998)
20. P.D. Ditlevsen, I.A. Mogensen, Cascades and statistical equilibrium in shell models of turbulence, *Phys. Rev. E* **53**, 4785 (1996)
21. M.H. Jensen, G. Paladin, A. Vulpiani, Intermittency in a cascade model for 3-dimensional turbulence, *Phys. Rev. A* **43**, 798 (1991)
22. R. Dewar, Information theory explanation of the fluctuation theorem maximum entropy production and self-organized criticality in non-equilibrium stationary states, *J. Phys.* **A36**, 631 (2003)
23. R. Dewar, Maximum entropy production and the fluctuation theorem, *J. Phys.* **A38**, L371 (2005)
24. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
25. L. Onsager, S. Machlup, Fluctuations and irreversible processes, *Phys. Rev.* **91**, 1505 (1953)
26. S.K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985)
27. D.J. Evans, D.J. Searles, L. Rondoni, Application of the Gallavotti-Cohen fluctuation relation to thermostated steady states near equilibrium, *Phys. Rev. E* **71**, 056120 (2005)
28. L. Rondoni, C. Mejà-Monasterio, Fluctuations in nonequilibrium statistical mechanics: models, mathematical theory, physical mechanisms, *Nonlinearity* **20**, R1 (2007)
29. U. Marini Bettolo Marconi, A. Puglisi, L. Rondoni, A. Vulpiani, Fluctuation-dissipation: Response theory in statistical physics, *Phys. Rep.* **461**, 111 (2008)
30. E.G.D. Cohen, G. Gallavotti, Note on two theorems in nonequilibrium statistical mechanics, *J. Stat. Phys.* **96**, 1343 (1999)

31. D.J. Searles, L. Rondoni, D.J. Evans, The steady state fluctuation relation for the dissipation function, *J. Stat. Phys.* **128**, 1337 (2007)
32. F. Bonetto, G. Gallavotti, A. Giuliani, F. Zamponi, Chaotic hypothesis, fluctuation theorem and singularities, *J. Stat. Phys.* **123**, 39 (2006)
33. F. Bonetto, G. Gallavotti, Reversibility, coarse graining and the chaoticity principle, *Commun. Math. Phys.* **189**, 263 (1997)
34. L. Rondoni, G.P. Morriss, Large fluctuations and axiom-Cstructures in deterministically thermostatted systems, *Open Syst. Inf. Dynam.* **10**, 105 (2003)
35. G. Gallavotti, L. Rondoni, E. Segre, Lyapunov spectra and nonequilibrium ensembles equivalence in 2D fluid mechanics, *Physica D* **187**, 338 (2004)
36. J. Farago, Injected power fluctuations in Langevin equation, *J. Stat. Phys.* **107**, 781 (2002)
37. G.M. Wang, E.M. Sevick, E. Mittag, D.J. Searles, D.J. Evans, Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales, *Phys. Rev. Lett.* **89**, 050601 (2002)
38. R. van Zon, E.G.D. Cohen, Stationary and transient work-fluctuation theorems for a dragged Brownian particle, *Phys. Rev. E* **67**, 046102 (2003)
39. M. Baiesi, T. Jacobs, C. Maes, N.S. Skantzos, Fluctuation symmetries for work and heat, *Phys. Rev. E* **74**, 021111 (2006)
40. T. Mai, A. Dhar, Nonequilibrium work fluctuations for oscillators in non-Markovian baths, *Phys. Rev. E* **75**, 061101 (2007)
41. R.J. Harris, Rákos, M. Schütz, Breakdown of Gallavotti-Cohen symmetry for stochastic dynamics, *Europhys. Lett.* **75**, 227 (2006)
42. L. Conti, P. De Gregorio, G. Karapetyan, C. Lazzaro, M. Pegoraro, M. Bonaldi, L. Rondoni, *Effects of breaking vibrational energy equipartition on measurements of temperature in macroscopic oscillators subject to heat flux*, *JSTAT* **2013**, P12003 (2013)
43. G. Auletta, A Paradigm Shift in Biology? *Information* **1**, 28 (2010)
44. G. Auletta, *Cognitive Biology: Dealing with Information from Bacteria to Minds*, (Oxford University Press, 2011)
45. I. Prigogine, *Etude thermodynamique des phénomènes irréversibles* (Desoer, Liège, 1947)
46. K.J. Friston, B. Sengupta, G. Auletta, Cognitive Dynamics: From Attractors to Active Inference, *Proc. IEEE* **102**, 427 (2014)
47. Ch. Feinauer, M.J. Skwark, A. Pagnani, E. Aurell, Improving Contact Prediction along Three Dimensions, *PLOS Comp. Biol.* **10**, e1003847 (2014)