PAPER • OPEN ACCESS

Measures of thermodynamic irreversibility in deterministic and stochastic dynamics

To cite this article: Ian J Ford 2015 *New J. Phys.* **17** 075017

IOP Publishing New J. Phys. 17 (2015) 075017 doi:10.1088/1367-2630/17/7/075017

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

RECEIVED
20 April 2015

REVISED
11 June 2015

ACCEPTED FOR PUBLICATION
23 June 2015

PUBLISHED
3 August 2015

PAPER

Measures of thermodynamic irreversibility in deterministic and stochastic dynamics

Ian J Ford

Department of Physics and Astronomy and London Centre for Nanotechnology, University College London, Gower Street, London, WC1E 6BT, UK

E-mail: i.ford@ucl.ac.uk

Keywords: irreversibility, entropy, dissipation function, Past Hypothesis, fluctuation relations

Abstract

It is generally observed that if a dynamical system is sufficiently complex, then as time progresses it will share out energy and other properties amongst its component parts to eliminate any initial imbalances, retaining only fluctuations. This is known as energy dissipation and it is closely associated with the concept of thermodynamic irreversibility, measured by the increase in entropy according to the second law. It is of interest to quantify such behaviour from a dynamical rather than a thermodynamic perspective and to this end stochastic entropy production and the time-integrated dissipation function have been introduced as analogous measures of irreversibility, principally for stochastic and deterministic dynamics, respectively. We seek to compare these measures. First we modify the dissipation function to allow it to measure irreversibility in situations where the initial probability density function (pdf) of the system is asymmetric as well as symmetric in velocity. We propose that it tests for failure of what we call the obversibility of the system, to be contrasted with reversibility, the failure of which is assessed by stochastic entropy production. We note that the essential difference between stochastic entropy production and the time-integrated modified dissipation function lies in the sequence of procedures undertaken in the associated tests of irreversibility. We argue that an assumed symmetry of the initial pdf with respect to velocity inversion (within a framework of deterministic dynamics) can be incompatible with the Past Hypothesis, according to which there should be a statistical distinction between the behaviour of certain properties of an isolated system as it evolves into the far future and the remote past. Imposing symmetry on a velocity distribution is acceptable for many applications of statistical physics, but can introduce difficulties when discussing irreversible behaviour.

1. Introduction

Thermodynamic irreversibility refers to the fact that most physical processes that take place in the world tend to evolve in a preferred direction as time goes by. Common examples include the cooling of a hot object towards the temperature of its environment, the diffusion of coloured ink into an initially clear body of water, and on a broader scale, the aging of the body, the crumbling of mountains and the explosive death of stars. These events are never seen to proceed in the opposite direction. Before the development of theories of such phenomena on the microscopic scale, the only explanation that could be offered for such behaviour was the second law of thermodynamics, an empirical rule that the thermodynamic entropy of a system and its environment should never decrease as time progresses [1, 2].

Entropy has a reputation for being difficult to pin down, but it is sufficient at this stage to regard it simply as a physical property of objects that can be measured from heat flows and temperature changes in carefully conducted thermal experiments [3]. The entropy of the world can be regarded as a sum of entropies of its component parts and this sum never decreases as thermomechanical processes take place: the production of entropy seems to provide a measure of the irreversibility of this behaviour, the extent to which a process cannot be undone. To paraphrase Orwell, all processes are irreversible, but some are more irreversible than others [4]. For example, a process where mechanical work is performed rapidly on a thermostated system can be regarded as more irreversible than a slower process since it typically leads to the generation of a greater amount of entropy in the system and its environment after a given final state of the system is reached. In the limit where the process takes place over an infinite time, or quasistatically (also referred to as the thermodynamically *reversible* limit) the amount of generated entropy is zero.

As is widely known, there is a deep problem of compatibility between the second law and the underlying time reversal symmetric dynamics that we believe govern the evolution of the world [5–13]. The dynamics are consistent with scenarios where heat spontaneously flows from cold into warm bodies, and mountains uncrumble. That such events are not seen suggests that the initial conditions for such a history simply have never been encountered. By implication, the associated earlier state of the world that would give rise to such conditions did not arise, and the argument can be continued, in principle, back to the beginning of the Universe. Explanations of irreversibility can involve, then, either breaking the time reversal symmetry of the dynamics or by placing a constraint on the configurations that might be assumed by a system, equivalent to prescribing an initial condition of the Universe. The latter explanation is known as the Past Hypothesis [9, 14].

Much has been written on these topics. The main subject of this paper, however, is to consider the status of two recently developed measures of irreversibility in classical dynamical systems: quantities that evolve in a fashion that resembles thermodynamic entropy when what seems to be irreversible behaviour takes place. The quantities in question are the integrated dissipation function, introduced in the context of deterministic system dynamics by Evans and collaborators [15–21], and stochastic entropy production, proposed for systems described by stochastic dynamics [22–26]. We discuss how they stand in relation to the breakage of time reversal symmetry in the dynamics and the Past Hypothesis, and how they compare with measures of irreversibility derived from Boltzmann's insights [27]. We seek conceptual similarities between the two measures in order to demonstrate more clearly their differences. We explore how they match up to the requirements of an analogue of thermodynamic entropy, as well as how they compare with ideas of pattern formation and dissolution. The Past Hypothesis seems the most acceptable way to rationalise irreversible behaviour (indeed it is almost a tautology) but it can imply certain modifications of the fundamental postulate of statistical physics in some circumstances, and this can have implications with regard to measures of irreversibility in a deterministic framework.

In the next section, we discuss the idea of a measure of irreversibility in more detail, and briefly review Boltzmann's phase space arguments as well as related issues of pattern formation and loss. In section 3 we discuss Loschmidt's objections to Boltzmann's work and describe the implications of the Past Hypothesis using a lighthearted illustration. Readers familiar with this material might go directly to section 4 where we define a modified version of the dissipation function (extending its remit to velocity asymmetric initial conditions) and interpret its time integral as a test of what we call the *obversibility* of the dynamics. In section 5 we contrast this quantity with stochastic entropy production which we present as a test of the subtly distinct *reversibility* of the dynamics. Both are assessed as measures of irreversibility. In section 6 we review the mathematical properties of the modified dissipation function, particularly for velocity asymmetric cases (with further results given in the appendix), and in section 7 we illustrate these properties using a simple dynamical system. Incorporating velocity asymmetry into the fundamental postulate of statistical physics, as well as the rationale for not doing so, is discussed in section 8, and in section 9 we give our conclusions.

2. Measures of irreversibility

The UK television company Channel 4 currently runs some very distinctive idents: short sequences that are broadcast just before the start of a programme to identify the channel and to contribute to the branding. The idents in question show objects, very often buildings or road signs, from the point of view of a moving observer. The key moment in the sequence is when the objects come together to form the shape of the number 4, before separating once again into a relatively meaningless pattern. They can be very imaginative [28].

A similar sequence of images showing the formation and dissolution of the letter *S* is shown in figure 1. Though they might seem to have little relevance to a discussion of thermodynamic irreversibility, the idents demonstrate some problems that emerge when we consider the evolution of order and disorder in such a context. For while the two frames on the right in figure 1 seem to show the progressive emergence of disorder with respect to the pattern apparent in the middle frame, this is not necessarily a measure of irreversible behaviour. Clearly, the pattern in question is generated fleetingly in the sequence, as illustrated by the two frames on the left. If we take the frames to be a sequence of snapshots in time, then disorder decreases from the past to the present and then increases again; in other words there is a loss of the coherence of a pattern towards the past as well as the future. Indeed there is nothing to suggest that the time order in the sequence runs from left to right or right to left: a statement of the time reversal symmetry of the dynamics. Furthermore, the pattern that forms

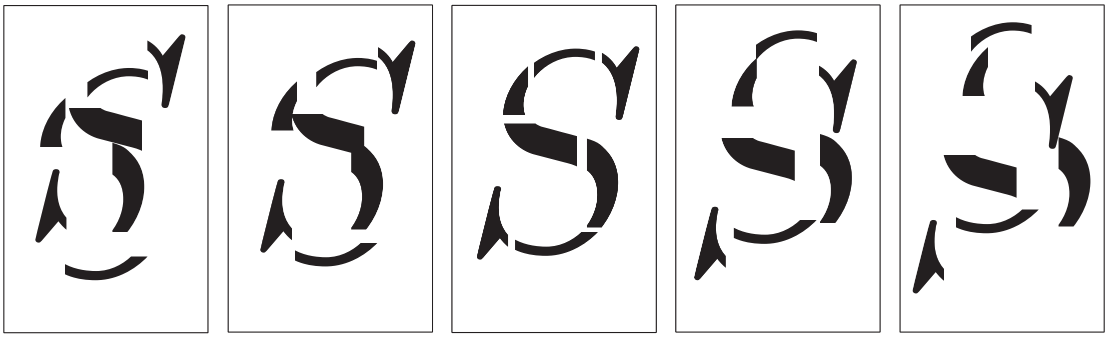

Figure 1. The letter *S* forms from moving objects and dissolves again as they continue to migrate (the sequence might be considered to evolve from left to right or the other way round). This can illustrate several problems in establishing a dynamical understanding of thermodynamic irreversibility.

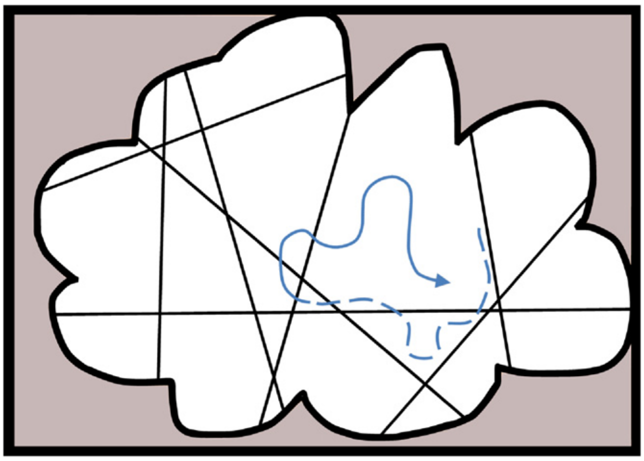

Figure 2. Boltzmann's conception of a dynamical underpinning of the second law. Accessible phase space is represented by a white region and is subdivided into patches of varying size according to selected collective system properties. A system that takes coordinates at $t = 0$ somewhere inside a small patch will most likely evolve away from the patch and as $t \to \infty$ is likely to be found in the largest patch available, as shown by the solid trajectory terminated by an arrowhead. However, the argument would seem to work equally well as $t \to -\infty$ (dashed trajectory).

and dissolves depends on the perception of the observer: why should the central frame be regarded as ordered at all? It seems more the case that figure 1 illustrates the *passage* of time rather than a directionality; at most it shows a loss in perceived order, with respect to an arbitrarily chosen starting configuration, that is symmetric in evolution towards the past and the future. Similar problems are to be found in more serious attempts to understand irreversibility.

Consider, for example, the explanation of the second law based on Boltzmann's well known representation of entropy, namely that it is a measure of the number of ways in which a system might be rearranged on a microscopic scale without changing its macroscopic features. We imagine setting up a system with initial macroscopic features that correspond to a small subset of the phase space. The system can take any one of the configurations within this patch. If the dynamics are complex, it is argued that as time progresses the system will leave behind the patch of phase space where its evolution began, virtually never to return [10, 29–31]. This would come about because the number of configurations to explore beyond this patch is immensely large and the dynamics extremely rich. Macroscopic features will evolve accordingly, and will only stop changing when the system reaches a patch of phase space so vast in size that the likelihood of a move into a smaller patch is negligible. The size of the patch corresponds to the Boltzmann entropy of the system and so a typical evolution of a system will be accompanied by an increase in this quantity. This is illustrated in figure 2. The evolution of the Boltzmann entropy with time is a measure of progress towards the equilibrium state, and of the irreversibility of the process, since going back to a previous patch becomes more unlikely as the current size of patch increases.

But an obvious problem is apparent when we imagine the behaviour of the system before it started its trajectory out of the small patch of phase space. If we use the same dynamical rules and simply run the clock backwards, the system will surely wander in a similar fashion into the wider phase space. There would be an increase in Boltzmann entropy into both the past and the future starting from the chosen initial conditions. Boltzmann was aware of this 'minimum entropy problem'. The implication is that the starting condition was merely a fluctuation; a fleeting manifestation of order (in some sense) preceded and succeeded by disorder. Demonstrations of such behaviour were obtained in some of the earliest numerical simulations of particle dynamics [32, 33]. This is not what we require: we wish the backward trajectory of the system to gravitate towards ever *smaller* patches of phase space. The situation is analogous to the Channel 4 idents and the sequence in figure 1, and cannot without further development be a solution to the problem of irreversibility.

Indeed it was pointed out by Boltzmann's colleague Loschmidt that if the dynamics and the statistics of the initial state respected time reversal symmetry, then evolution into the past ought to be indistinguishable from evolution into the future [27]. Boltzmann could offer a kinetic model of irreversibility (the celebrated *H*-theorem) that evaded Loschmidt's objection by introducing a time asymmetry into the dynamics, albeit in a subtle fashion. It is apparent, however, that another solution would be to demand asymmetry in the statistics, which can amount to constraints placed on system configurations at remote times. In the next section we consider how such boundary conditions allow us to evade Loschmidt's objection.

3. Evading Loschmidt through the Past Hypothesis

If the dynamics are symmetric under time reversal, we can reconstruct events that took place prior to an arbitrary starting configuration of a system by using a configuration of the system with inverted velocities as the starting point instead. This would include an inversion of any other odd parity variables as well. As time progresses forward, this configuration evolves through the (velocity-inverted) configurations previously adopted by the system.

The procedure is illustrated in figure 3, in a lighthearted way, using trains on a railway network (an example of deterministic dynamics, allegedly). The situation at the top is denoted configuration 1 with train positions and directions of motion as shown. We can imagine that this evolves into configuration 2 as time progresses forward. In order to retrace the trajectory that took place prior to configuration 1, we would invert the initial velocities and allow time to progress forward starting from configuration 3. We would then see the rearward moving trains adopting positions occupied prior to configuration 1. This much is clear.

Loschmidt's argument can now be framed as follows. Boltzmann's phase space method involves determining the likely behaviour of a system given that the initial configuration is a member of a certain collective or patch of phase space (a macrostate). Various statistical weights might be assigned to each member of this collective, over which an average is to be taken in order to deduce what is likely to happen. But if both configurations 1 and 3 are possible initial configurations (with the same weight), then whatever expectation we might gather from the evolution of one would be cancelled out by the evolution of the other.

Applying such logic would challenge the picture in figure 2: for each point in the initial patch of phase space that is on the verge of entering a larger one there would be another leaving the larger one to enter the smaller. Only an expectation of static behaviour would emerge from such a collective of configurations.

The key assumption, of course, is that both configuration 1 and 3 are equally weighted members of the initial collective, and here we see how Loschmidt's objection can be evaded. Perhaps velocity-inverted configurations should not both be included in the set of possible starting configurations? Less drastically, perhaps configurations 1 and 3 might be accorded unequal weights. Both amount to breaking the symmetry of the initial statistics under velocity inversion.

In fact, there could be very good physical reasons for configuration 3 to be discarded as a possible starting arrangement if configuration 1 is allowed, and vice versa. Suppose, for example, that the trains were set moving at the some point in the remote past with the rule that they should thereafter only be allowed to move *forward*? Then at arbitrary times in the evolution, the collective of possible configurations will only contain trains that are moving in that direction. Configuration 3 would not be a member of this collective. This would be a version of the Past Hypothesis. The rule that trains are only allowed to move forward would apply to the system whatever moment in time is of interest, and it would be a reflection of a condition imposed on the dynamics at the very earliest time under consideration. There would be parts of phase space that would forever be inaccessible.

The phase space explanation of irreversibility shown in figure 2 can then be made to work. If the patch of phase space from which the trajectory emanates does not contain equally weighted velocity-inverted pairs of configurations, then the likely evolution into the past will not resemble an evolution into the future and it is possible that the system might on average gravitate back in time towards ever smaller regions of phase space.

Might collectives of configurations of *particles* be constrained in this way? Should we include velocity-inverted pairs of configurations or does the inclusion of one exclude the other? This depends very much on how the system is prepared, and which properties we are interested in, which is to be discussed in section 8. Giving both configurations equal weight when judging future behaviour would make it difficult to explain irreversible

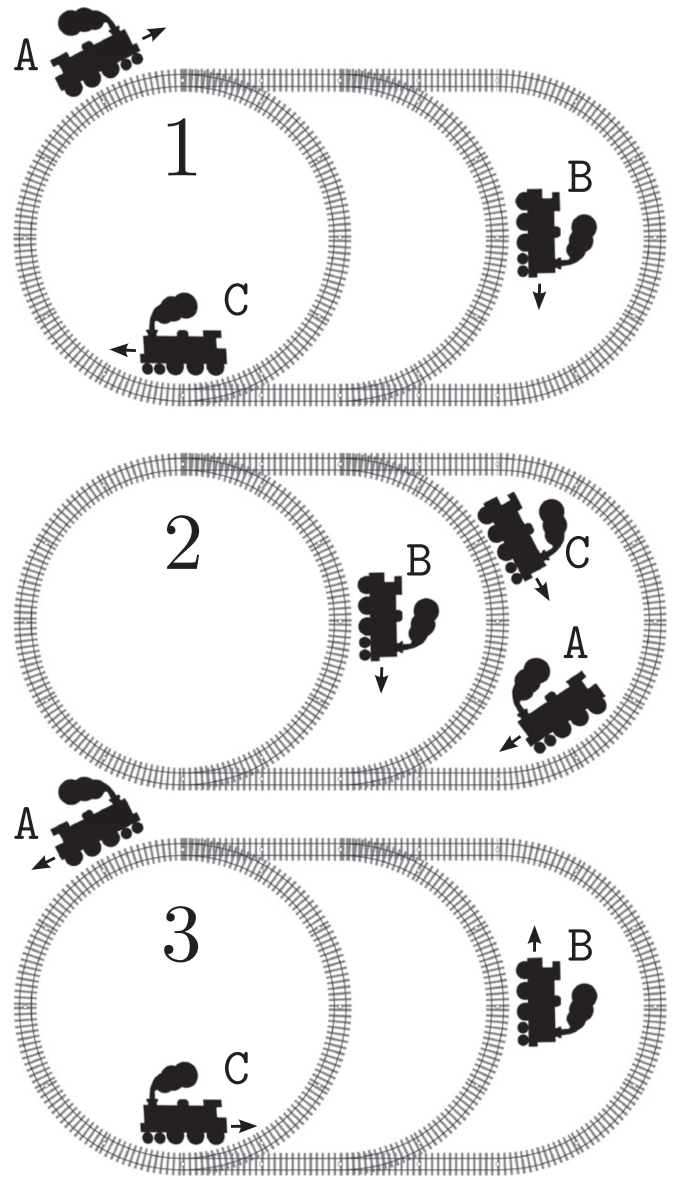

Figure 3. We consider the dynamics of three trains, A, B and C, with initial positions and directions of motion on a track shown in configuration 1. After time has progressed, the trains assume configuration 2. In order to determine the configurations taken prior to configuration 1, we would invert velocities to form configuration 3, and progress forward in time from there.

phenomena such as relaxation. We shall bear this in mind in the next section, where we discuss an irreversibility measure for systems that evolve under deterministic dynamics.

4. Measuring irreversibility in deterministic dynamics

Readers well-versed in the arguments made by Boltzmann and Loschmidt, and familiar with the Past Hypothesis, will have been able to skip quickly through the previous two sections. The main purpose of this paper is to compare measures of irreversibility that have been developed for the classical dynamical frameworks of deterministic and stochastic equations of motion. We begin with deterministic rules.

We should start by carefully considering what is meant by *reversibility*. If we evolve a system along a trajectory for a period of time starting from an arbitrary initial configuration, and then invert velocities and evolve for the same period under a reverse sequence of external forces, reversibility is assessed by determining whether the previous sequence of configurations is retraced (with inverted velocities); that is whether the system then follows an *antitrajectory*. But since we are assuming that the dynamics respect time reversal symmetry, this procedure will produce the antitrajectory every time. As an example, consider how configuration 1 in figure 3 evolves into configuration 2, which might then be converted into configuration 4 in figure 4 by an inversion of velocities, and how as time progresses further still, the trains will return to their original positions, albeit with direction of

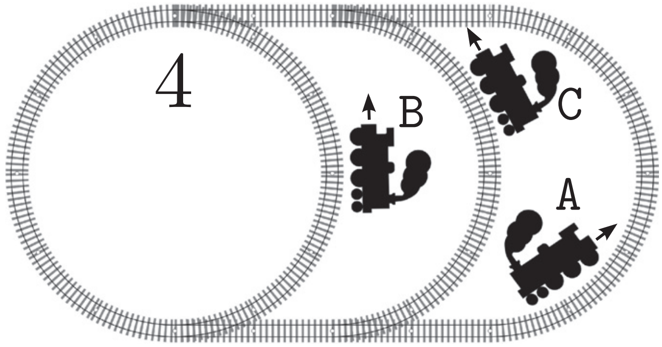

Figure 4. In order to generate a sequence of events corresponding to the reverse of the trajectory taking configuration 1 into configuration 2 in figure 3, we need to start with configuration 4, corresponding to configuration 2 with inverted velocities, and then proceed forward in time until the system reaches configuration 3. Note that configuration 4 would not be a member of the collective of possible initial configurations if the collective were subject to the rule that trains could only move forward, in adherence to a form of the Past Hypothesis.

motion to their rear, in other words configuration 3. This will not provide a measure of irreversibility since such an outcome is inevitable.

But there is another approach, developed by Evans and collaborators [15, 16]. We represent uncertainty in initial configuration using a probability density function (pdf). For a given trajectory and protocol of external forces, we use the pdf to gauge the likelihood that amongst all the possible *initial* configurations of a system there might be one that would produce the associated antitrajectory. The system would be indifferent to the directionality of time if the likelihoods of following a trajectory and corresponding antitrajectory were equal, and a difference between these probabilities might form the basis of a measure of irreversibility. It must be possible to find the initial configuration of the antitrajectory amongst the possible initial configurations of the system, a condition known as ergodic consistency, and also the protocol should either be time independent or time symmetric over the period in question.

For example in our trains analogy, we compare the probabilities that the system should initially take configuration 1 or configuration 4. A disparity between these probabilities will lead to an expectation that movement should take place one way rather than the other, and a directionality of evolution might emerge.

However, there is a problem with this approach. If we consider this comparison in the limit of a process of zero duration, we find that the probabilities in question are those initially attached to configurations 1 and 3. We do not expect to measure any irreversibility for a process of zero duration, so these probabilities had better be equal. The problem is that we argued in the last section that this is potentially incompatible with a Past Hypothesis. An extreme form of this difficulty is where configurations 4 and 3 simply are unavailable as starting configurations as a consequence of a rule on the allowed directionality of train motion. This would then reduce to the breakage of the ergodic consistency requirement. An asymmetry in probabilities of velocity-inverted configurations is different, but it represents a similar problem.

So we should modify our assessment procedure. Let us instead compare the probability of a trajectory with the probability of following an antitrajectory after selecting an initial condition *and then inverting velocities* before proceeding forward in time. We might furthermore imagine that the subsequent evolution takes place under a time reversed protocol of forces, if applicable. The starting configuration for the antitrajectory would be configuration 2: converting it to configuration 4 would be part of the test procedure. Considering a process of zero duration does not now present a problem, since the probabilities for the two events would tend to the same value, namely the initial probability of configuration 1. We do not require the initial pdf to be symmetric in velocity. We would not seek an initial configuration with rearward moving trains only to find that none were available. By considering the difference in the probabilities of generating trajectories and antitrajectories in this way, we could establish a measure of the irreversibility. This is a small but important modification of the procedure introduced by Evans and collaborators, where velocity inversion is not carried out after the selection of a configuration. The standard and modified procedures are illustrated in figure 5.

It might be best to describe such an assessment of irreversibility using new terminology to reflect the nature of this antitrajectory generation procedure. We are not assessing how we might *reverse* a trajectory. That idea is naturally associated with the procedure of inverting velocities at the *end* of a trajectory. We are instead assessing the production of both trajectory and antitrajectory according to the *initial* configurational statistics of the system.

This key point is worth repeating. The words *reversing* or *reversibility* refer to comparing the behaviour of a system over a period of time with its behaviour over a *subsequent* period of time, when a procedure is followed to

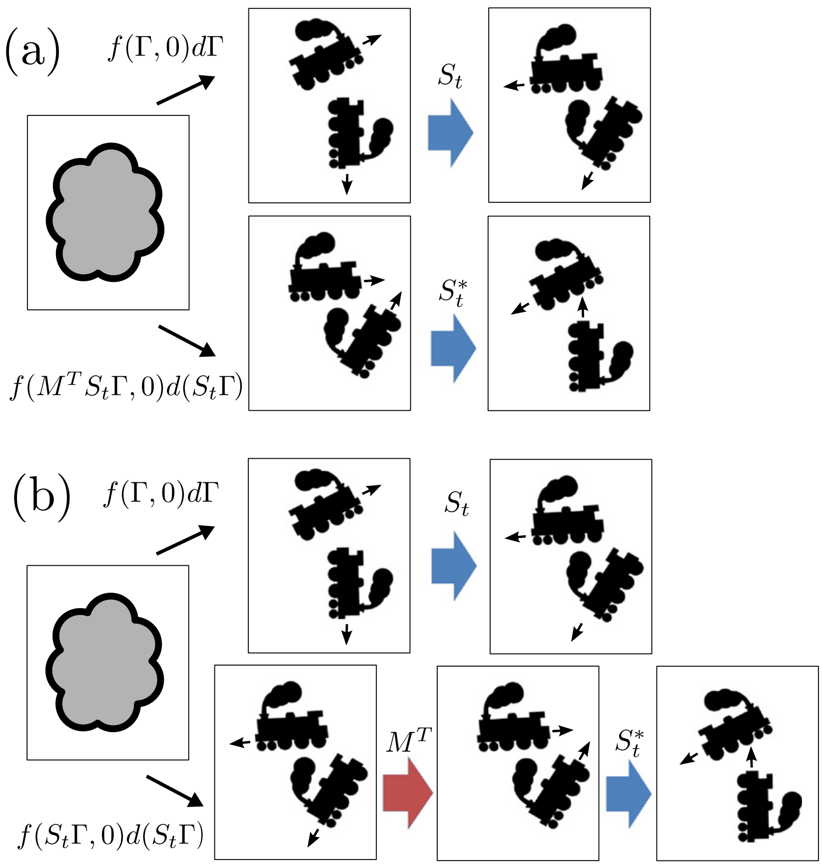

Figure 5. Illustration of procedures used to measure irreversibility in deterministic dynamics. The initial configuration adopted by a system is uncertain (represented by a cloud). In (a), we follow Evans and collaborators and compare the probabilities that the initial configuration might be Γ or $M^T S_t \Gamma$, where S_t represents evolution through a time t under a certain protocol and M^T is a velocity inversion. Configurations Γ and $M^T S_t \Gamma$ lie at the start of a trajectory and corresponding antitrajectory, respectively, the latter evolution taking place under a reverse protocol and represented by S_t^*. Note that if the protocol is symmetric about the mid-point of the period we would have $S_t = S_t^*$. The probabilities are given as a product of a pdf f and an increment in phase space. In (b), we compare instead the probabilities that the initial configuration might be Γ or $S_t \Gamma$: the latter is the configuration that, *after* a velocity inversion, would produce the antitrajectory. We call (b) a test of *obversibility*. Selection is denoted by black arrows, time evolution by blue, and velocity inversion by red. For stochastic dynamics the procedures would be similar except that the time evolution would not be a unique mapping of initial to final configuration.

try to undo the change and bring the world back to where it started. The antitrajectory is to be generated after a certain forward process has been completed. We assess reversibility through the statistics of producing the antitrajectory by velocity inversion and applying a reversed protocol *at a point in the future*. If the probability of producing an antitrajectory is equal to that of producing the associated trajectory in the preceding interval of time, then the dynamics would be said to be *reversible* for this trajectory. Complete reversibility means the equality holds for an arbitrary trajectory. The terminology ties in with what is meant by time reversal symmetry or reversibility in deterministic dynamics. So the extent to which the system can generate trajectories and *subsequent* antitrajectories with equal likelihood is its reversibility. The failure of complete reversibility is a measure of irreversibility.

In contrast, when we assess how we might produce a trajectory or its antitrajectory using the statistics of initial configurations, we are testing for something else, and a word that comes to mind is 'obversibility'. An *obverse* of an object or idea is in some sense the object 'turned round'. The dynamics and initial pdf of the system would be completely 'obversible' if an arbitrary trajectory and its associated antitrajectory (obtained after velocity inversion and evolving under a time reversed protocol) can be produced from the *start* of the period in question with equal probabilities. The extent to which the system can achieve this could be referred to as its obversibility and would depend on the difference in probabilities of generating trajectories and antitrajectories. The failure of complete obversibility is a measure of irreversibility quite distinct from the failure of complete reversibility.

For a stationary situation, where the pdf describing the coordinates of a system is time independent, the pdf of final configurations of the system is the same as that for initial configurations and the outcomes of the tests of reversibility and obversibility must coincide. Away from stationarity, the two measures of irreversibility differ from one another in general.

In systems that evolve deterministically under dynamics that respect time reversal symmetry, reversibility is automatic, and we recognised earlier that such an assessment cannot provide a measure of irreversibility. So for deterministic dynamics we must employ an obversibility test to seek a measure of irreversibility. Up to a point, a suitable quantitative procedure exists already.

Evans and collaborators introduced the concept of the dissipation function in order to assess irreversibility. The time integral of the dissipation function is defined as the logarithm of the ratio of the probabilities of producing a trajectory and an antitrajectory for a given interval starting from the same moment in time. For illustration, let us imagine that a trajectory runs from phase space coordinates (x_1, v_1) at time t_1 to (x_2, v_2) at time $t_2 \geqslant t_1$, and an antitrajectory runs from $(x_2, -v_2)$ to $(x_1, -v_1)$ over the same period. In terms of our trains analogy, this would correspond to the trajectory running from configuration 1 to 2, and the antitrajectory running from configuration 4 to 3. The time-integrated dissipation function is defined in terms of a ratio of the initial probability densities at phase space coordinates (x_1, v_1) and $(x_2, -v_2)$, together with a phase space expansion/contraction factor that compares the size of a patch of phase space surrounding (x_1, v_1) at time t_1 to the size of a patch around (x_2, v_2) at time t_2 that represents the time evolved initial patch. Details may be found in the literature [15, 16].

With reference to figure 5(a) the dissipation function Ω and its time integral Ω_t are defined as

$$\Omega_t(\Gamma) = \int_0^t \Omega\left(S_s\Gamma\right)\mathrm{d}s = \ln\left[\frac{f(\Gamma, 0)\,\mathrm{d}\Gamma}{f\left(M^T S_t\Gamma, 0\right)\mathrm{d}\left(S_t\Gamma\right)}\right], \tag{1}$$

where $f(\Gamma, t)$ is the pdf of coordinates Γ at time t and S_s is an operator that evolves coordinates through time s according to the dynamics. M^T is a velocity inversion operator that changes the sign of odd variables such as velocity. It ought also to appear inside the phase space increment in the denominator on the right hand side of equation (1), but we note that velocity inversion has a Jacobian of unity and so it can be ignored.

It should be noted that Evans and collaborators assume the initial pdf $f(\Gamma, 0)$ to be symmetric in velocity coordinates, such that the initial probability density at coordinates $(x_2, -v_2)$ is equal to the initial probability density at coordinates (x_2, v_2), or $f(M^T S_t\Gamma, 0) = f(S_t\Gamma, 0)$. The integrated dissipation function may therefore in effect be written in terms of the initial pdf evaluated at (x_1, v_1) and (x_2, v_2). Assuming ergodic consistency, the pdf at both these phase space points would be non-zero. Furthermore, as the duration of the process goes to zero, these coordinates merge, the probability densities and phase space increments inside the logarithm in equation (1) become equal and the integrated dissipation function vanishes.

As mentioned earlier, the protocols of external forces considered using this procedure are often taken to be time independent or time symmetric about the process mid-point. The integrated dissipation function was clearly designed to compare the probabilities of generating trajectories and antitrajectories *without* in the latter case inverting velocities immediately and reversing the time sequence of forces. The symmetries assumed make the additional procedures unnecessary. It is an assessment of the likelihood of generating trajectory and antitrajectory 'within the same ensemble', in a manner of speaking.

Our modified procedure for measuring obversibility similarly requires us to compute the logarithm of the ratio of probabilities of producing a trajectory and the antitrajectory, but the latter now consists of the process of selecting a configuration from the initial pdf, inverting velocities and evolving under a reversed protocol, (see figure 5(b)). This now concerns the initial probability density at phase space coordinates (x_2, v_2), not $(x_2, -v_2)$, and our measure of obversibility is therefore a function of the initial pdf evaluated at (x_1, v_1) and (x_2, v_2). The logarithm on the right hand side of equation (1) would be replaced by $\ln[f(\Gamma, 0)\mathrm{d}\Gamma / f(S_t\Gamma, 0)\mathrm{d}(S_t\Gamma)]$ and this would define a modified dissipation function and its time integral. To align the mathematics to the procedure, we ought to append a probability $P_I(S_t\Gamma \to M^T S_t\Gamma)$ to the denominator to represent the conversion of $S_t\Gamma$ to $M^T S_t\Gamma$, corresponding to the red arrow in figure 5, but this would be unity. We emphasise that the integrated dissipation function used by Evans and collaborators is also a function of the initial pdf at (x_1, v_1) and (x_2, v_2), but only by virtue of the assumed symmetry in velocity.

We can use the modified dissipation function as a measure of obversibility according to our test procedure, and the different perspective means that we shall not be restricted to systems possessing a velocity symmetric initial pdf and evolving under a time symmetric protocol. We do not assess whether a trajectory and antitrajectory might be generated 'within the same ensemble' since our assessment of the latter also includes the application of a velocity inversion and a reverse protocol.

We shall explore the mathematics of the modified dissipation function in the context of velocity asymmetric initial pdfs in section 6. First, we note that although reversibility is automatic in systems governed by time reversal symmetric deterministic dynamics, such that a test for it cannot provide a measure of irreversibility, this is not the case in systems evolving by stochastic dynamics. We look at how tests of both obversibility and reversibility can provide measures of irreversibility for stochastic systems in the next section.

5. Measuring irreversibility in stochastic dynamics

In deterministic classical dynamics, the gain or loss of heat by a system during its evolution is represented by inserting nonlinear terms in the equations of motion such that elementary phase space volumes are not preserved. These terms represent the action of a thermostat or thermal environment. In their absence Liouville's theorem would imply that the Gibbs entropy of the system is conserved, a commonly raised objection to the use of Hamiltonian dynamics as a framework for understanding irreversibility. In stochastic dynamics, by contrast, the effects of the environment are represented by a noise term, and this destroys the automatic reversibility of the dynamics. In addition, deterministic terms that explicitly break time reversal symmetry are often inserted.

It should be noted that stochastic equations of motion are an effective description of the behaviour after the neglect of some detail in the system or its environment. They can be referred to as macrostate, projected or coarse-grained dynamics, and the stochasticity is essentially a reflection of configurational uncertainty in the parts of the world that are not being considered in detail.

The stochastic evolution of a system has long been regarded as a suitable framework for discussing irreversibility. We shall focus our attention on the recently developed concept of stochastic entropy production, but a broader discussion extends back at least as far as Schnakenberg's [34] model of entropy production associated with a system of master equations. Tomé and de Oliveira have recently reviewed and developed this approach [35] where entropy production and flow are related to a system pdf and the rate coefficients that govern its evolution. The form of stochastic thermodynamics that we employ here, on the other hand, links these concepts to the noisy equations of motion of system dynamical variables instead of the deterministic evolution of a system pdf. A stochastic production of entropy emerges, which upon averaging over realisations of the noise will coincide with a pdf-based approach.

Under stochastic dynamics, a non-trivial assessment of reversibility, as conceived in section 4, is possible since an antitrajectory is not automatically generated by the procedure of inverting velocities and applying a reversed protocol of external forces starting at the end of a trajectory. The reason is that while the system or *projected* velocities are available for inversion, the velocity coordinates in the underlying dynamics that are not retained in the projection are not. Similarly, it is only the system velocities that are available for inversion at the beginning of a trajectory when we assess the obversibility. The situation can be illustrated by regarding the configurations in figures 3 and 4 as merely a small but visible part of a wider rail network, and that when we generate configuration 4 from configuration 2, we are not able to invert the velocities of the trains we cannot see, and thus configuration 4 might not necessarily evolve over time into configuration 3.

In order to assess the reversibility of a stochastic dynamical system we follow a procedure similar to that used to assess obversibility in deterministic dynamics, but with important differences. As before, we compute the logarithm of the probability of producing a trajectory driven by a protocol, divided by the probability of producing its antitrajectory. However, for the latter we select a configuration from the system pdf produced *after* a given period of time, and then invert velocities and apply a reversed protocol. This quantity has been investigated in the literature and is known as the stochastic entropy production [22–24, 26]. The key differences with respect to the integrated dissipation function are: the time at which the initial coordinates for the antitrajectory are selected; the absence, typically, of a phase space expansion/contraction factor associated with the deterministic terms in the dynamics since the action of a thermostat is represented by the noise terms; and the inclusion of probabilities associated with the stochastic evolution from initial to final coordinates.

The mathematical definition of stochastic entropy production is

$$\Delta s_{\rm tot}\left[\Gamma \to \Gamma'\right] = \ln\left[\frac{f(\Gamma, 0)\,T_t\left(\Gamma \to \Gamma'\right)}{f\left(\Gamma', t\right)P_I\left(\Gamma' \to M^T\Gamma'\right)T_t\left(M^T\Gamma' \to M^T\Gamma\right)}\right], \tag{2}$$

where Γ and Γ' are the initial and final coordinates of the trajectory and T_t is a probability density for the path taken between the two over a time interval t. Increments of phase space associated with the initial and final coordinates and the paths should also appear in numerator and denominator but they cancel. The key point is that the pdf in the denominator refers to configuration Γ' and not $M^T\Gamma'$. We have inserted a probability P_I for the inversion $\Gamma' \to M^T\Gamma'$ to align the mathematics with the test procedure, but this probability would be unity, and is typically not discussed in the literature. The procedure for assessing reversibility is illustrated in figure 6, and the contrast with respect to the procedures in figure 5, particularly option (b), should be noted.

Similarly, we can assess obversibility in a stochastic dynamical system by comparing the probability of producing a trajectory driven by a protocol, to the probability of producing its antitrajectory through selecting a

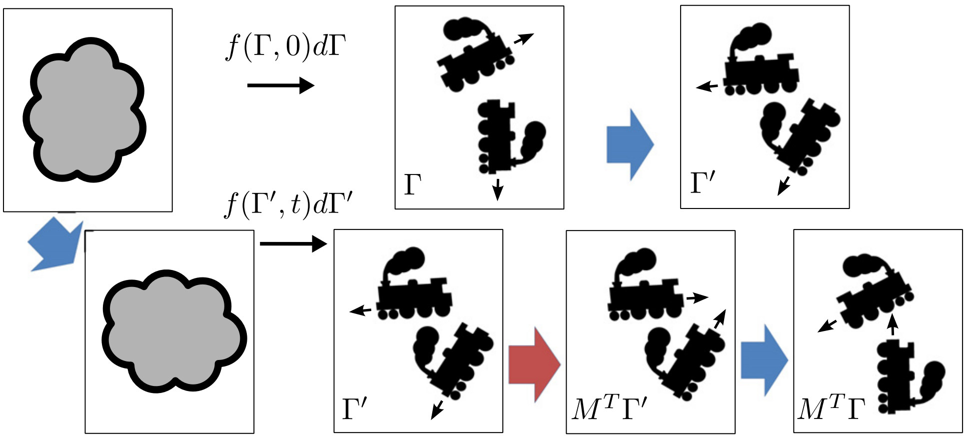

Figure 6. Illustration of the procedure used to assess reversibility. The initial configuration adopted by a system is denoted Γ. It is considered to evolve under a protocol, according to some probability, along a trajectory to assume configuration Γ' after time t (upper row of images) We also consider the probability that the system might adopt configuration Γ' after time t without specifying its previous behaviour (lower row). After a velocity inversion, denoted by a red arrow, the system could follow an antitrajectory for a further time t under a reversed protocol, again according to some probability. For time reversal symmetric deterministic dynamics, the probabilities of generating trajectory and antitrajectory in this way are the same and reversibility is automatic, but this is not so for stochastic dynamics.

configuration according to the *initial* system pdf, before inverting velocities and applying a reversed protocol. The logarithm on the right hand side of equation (2) would be replaced by $\ln[f(\Gamma, 0)T_t(\Gamma \to \Gamma')/f(\Gamma', 0)P_I(\Gamma' \to M^T\Gamma')T_t(M^T\Gamma' \to M^T\Gamma)]$. We might still call this an integrated dissipation function, but in contrast to equation (1) it involves the probabilities for paths as well as initial coordinates.

The relevance of stochastic entropy production to the second law and its status as a measure of irreversibility emerges from what is called, in this context, the integral fluctuation relation [22]. It is quite straightforward to demonstrate that the expected value of this quantity, taking into account the uncertainty in the initial condition and the stochastic dynamics of the process, is never negative. The proof is

$$\begin{aligned}\left\langle \exp\left(-\Delta s_{\rm tot}\right)\right\rangle &= \int {\rm d}\Gamma{\rm d}\Gamma' f(\Gamma, 0)T_t\left(\Gamma \to \Gamma'\right)\exp\left(-\Delta s_{\rm tot}\right),\\ &= \int {\rm d}\Gamma{\rm d}\Gamma' f\left(\Gamma', t\right)P_I T_t\left(M^T\Gamma' \to M^T\Gamma\right) = 1,\end{aligned} \tag{3}$$

having inserted $P_I = 1$, and this implies that $\langle \Delta s_{\rm tot}\rangle \geqslant 0$, since $\exp(-z) \geqslant 1 - z$ for $z \in \mathbb{R}$. Significantly, this holds for arbitrary initial statistics and process duration, so that the time evolution of $\langle \Delta s_{\rm tot}\rangle$ is monotonic. Numerous studies of stochastic dynamical systems have demonstrated that not only is this result reminiscent of the second law, but also that explicit evaluations of $\langle \Delta s_{\rm tot}\rangle$ can be shown to correspond to the production of thermodynamic entropy in various processes [25, 26].

Notice, however, that this outcome applies whether we run time backwards or forwards starting from the initial condition. It is usually the case that stochastic dynamics are introduced as a phenomenological description of evolution into the *future*, such that path probabilities for evolution into the past are not defined: the past is supposed to be known or recorded. This is not a requirement though. If we allow evolution into the past, but accept a Past Hypothesis, namely that retrodictions are expected to differ from predictions of behaviour in the future, then this could be accommodated by using a path probability density $T_{-t}(\Gamma \to \Gamma')$ that differs from $T_t(\Gamma \to \Gamma')$ [10]. The growth of $\langle \Delta s_{\rm tot}\rangle$ into the past would then be a reflection of an increasing uncertainty of retrodiction in circumstances where the past behaviour is not recorded, and would not then be the same as thermodynamic entropy production. Something akin to the pattern formation and dissolution in figure 1 would arise if we take the dynamics to operate in both directions in time starting from the initial condition, possibly with asymmetry. However, this does not spoil the demonstrated correspondence between the mean production of stochastic entropy and the generation of thermodynamic entropy for evolution into the future, and to reiterate, stochastic dynamics into the past can readily be regarded as inadmissible.

In the next section we turn our attention to the mathematical properties of the integrated modified dissipation function, with particular focus on those that rely on an assumption of velocity symmetry of the initial pdf, and enquire into similar points of past-future symmetry and asymmetry.

6. Properties of the integrated dissipation function for a velocity asymmetric initial pdf

6.1. Notation

Evans and his collaborators have explored the properties of the dissipation function while assuming that the pdf of initial configurations is symmetric in velocity, presumably motivated by a desire to investigate behaviour starting from a canonical equilibrium state. In order to make clear that our procedure for assessing obversibility does not require such symmetry, while it remains possible to use a very similar dissipation function, we shall use a different symbol. We define a modified dissipation function ω and its time integral through

$$\begin{aligned}\omega_t(\Gamma) &= \int_0^t \omega\left(S_s\Gamma\right)\mathrm{d}s,\\ &= \ln\left[\frac{f(\Gamma,\,0)\,\mathrm{d}\Gamma}{f\left(S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)P_I\left(S_t\Gamma\to M^TS_t\Gamma\right)}\right],\end{aligned} \tag{4}$$

where $f(\Gamma,\,0)$ is the pdf over phase space coordinates Γ at $t=0$. The dynamics take Γ into $S_s\Gamma$ after a time interval s. As before, note the difference between $\omega(\Gamma)$ and $\omega_t(\Gamma)$, and also the explicit inclusion of a probability P_I (equal to unity) associated with the velocity inversion of $S_t\Gamma$. It is the integrated modified dissipation function $\omega_t(\Gamma)$ that has the closest similarity to the stochastic entropy production defined in equation (2) and which will receive most of our attention. In order to proceed, we assume ergodic consistency in the form $f(S_t\Gamma,\,0)\neq 0$ if $f(\Gamma,\,0)\neq 0$.

Equation (4) should be contrasted with equation (1) for the dissipation function Ω given earlier:

$$\Omega_t(\Gamma) = \int_0^t \Omega\left(S_s\Gamma\right)\mathrm{d}s = \ln\left[\frac{f(\Gamma,\,0)\,\mathrm{d}\Gamma}{f\left(M^TS_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)}\right], \tag{5}$$

which is employed under an assumption that $f(M^T\Gamma,\,0)=f(\Gamma,\,0)$. Although the velocity symmetry condition is always clearly stated, definitions of Ω with and without the velocity inversion operator inside the pdf f in the denominator exist in the literature (for example comparing equation (2.6) in [16] with equation (1) in [19]) but here we explicitly include it.

Clearly ω_t might adhere to many of the relationships satisfied by Ω_t. Let us explore its properties.

6.2. Nonequilibrium partition identity (NPI)

We evaluate an expectation of the exponentiated integrated modified dissipation function weighted by the pdf at $t=0$, indicated by a suffix 0 on the angled brackets:

$$\begin{aligned}\left\langle\exp\left(-\omega_t(\Gamma)\right)\right\rangle_0 &= \int \mathrm{d}\Gamma\frac{f\left(S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)}{f(\Gamma,\,0)\,\mathrm{d}\Gamma}f(\Gamma,\,0),\\ &= \int f\left(S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right) = 1,\end{aligned} \tag{6}$$

where we have inserted $P_I=1$ (and will continue to do so from now on). This result is the counterpart of the NPI derived by Evans and collaborators:

$$\int \mathrm{d}\Gamma\,\exp\left(-\Omega_t(\Gamma)\right)f(\Gamma,\,0) = 1. \tag{7}$$

These results are analogues of the integral fluctuation relation satisfied by the stochastic entropy production in equation (3). The implication is that the mean integrated modified dissipation function is never negative:

$$\int \mathrm{d}\Gamma\omega_t(\Gamma)f(\Gamma,\,0) = \left\langle\omega_t(\Gamma)\right\rangle_0 \geqslant 0, \tag{8}$$

though its time development might not be monotonic.

6.3. Evans–Searles fluctuation theorem (ESFT)

A lack of velocity inversion symmetry in $f(\Gamma,\,0)$ does not spoil the NPI but it does impact another property. Using $M^TS_t^*M^TS_t\Gamma=\Gamma$, where S_t^* represents evolution under a reverse protocol, it may be shown that

$$\omega_t^*\left(M^T S_t \Gamma\right) = \ln\left(\frac{f\left(M^T S_t\Gamma,\,0\right)\mathrm{d}\left(M^T S_t\Gamma\right)}{f\left(S_t^* M^T S_t\Gamma,\,0\right)\mathrm{d}\left(S_t^* M^T S_t\Gamma\right)}\right),$$
$$= \ln\left(\frac{f\left(M^T S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)}{f\left(M^T\Gamma,\,0\right)\mathrm{d}\Gamma}\right),$$
$$= -\ln\left(\frac{f\left(M^T\Gamma,\,0\right)\mathrm{d}\Gamma}{f\left(M^T S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)}\right). \qquad (9)$$

The left hand side is the integrated modified dissipation function for an antitrajectory (we indicate the reverse protocol by the asterisk on ω_t) of points $\Gamma_s^* = S_s^* M^T S_t \Gamma$ (for $0 \leqslant s \leqslant t$) that arises from the velocity-inverted final configuration $M^T S_t \Gamma$ of the trajectory of points $\Gamma_s = S_s \Gamma$ (for $0 \leqslant s \leqslant t$) that evolves from phase space point Γ. Clearly, the modified dissipation function integrated along the antitrajectory is not, in general, equal and opposite in magnitude to the modified dissipation function integrated along the trajectory. This would hold, however, if $f\left(M^T\Gamma,\,0\right) = f\left(\Gamma,\,0\right)$ such that the right hand side of equation (9) reduces to the negative of the right hand side of equation (4), giving $\omega_t^*\left(M^T S_t\Gamma\right) = -\omega_t\left(\Gamma\right)$.

The implication of this is that the statistics of the integrated modified dissipation function do not necessarily possess a symmetry [16] known as the ESFT. The probability density function for ω_t is

$$P\left(\omega_t\right) = \int \mathrm{d}\Gamma f\left(\Gamma,\,0\right)\delta\left(\omega_t\left(\Gamma\right) - \omega_t\right),$$
$$= \int \mathrm{d}\Gamma_t f\left(\Gamma_t,\,t\right)\mathrm{e}^{\omega_t(\Gamma)}\frac{f\left(\Gamma_t,\,0\right)}{f\left(\Gamma_t,\,t\right)}\delta\left(\omega_t\left(\Gamma\right) - \omega_t\right),$$
$$= \mathrm{e}^{\omega_t}\int \mathrm{d}\Gamma_t f\left(\Gamma_t,\,0\right)\delta\left(\omega_t\left(\Gamma\right) - \omega_t\right), \qquad (10)$$

where we write $\Gamma_t = S_t\Gamma$ and have used the condition of conservation of probability $f\left(\Gamma,\,0\right)\mathrm{d}\Gamma = f\left(\Gamma_t,\,t\right)\mathrm{d}\Gamma_t$ such that $\omega_t\left(\Gamma\right) = \ln\left[f\left(\Gamma_t,\,t\right)/f\left(\Gamma_t,\,0\right)\right]$. If $f\left(M^T\Gamma,\,0\right) = f\left(\Gamma,\,0\right)$ and the protocol and reverse protocol are identical, such that $\omega_t\left(\Gamma\right) = -\omega_t\left(M^T\Gamma_t\right)$ we can write

$$P\left(\omega_t\right) = \mathrm{e}^{\omega_t}\int \mathrm{d}\Gamma_t f\left(\Gamma_t,\,0\right)\delta\left(-\omega_t\left(M^T\Gamma_t\right) - \omega_t\right),$$
$$= \mathrm{e}^{\omega_t}\int \mathrm{d}\left(M^T\Gamma_t\right) f\left(M^T\Gamma_t,\,0\right)\delta\left(\omega_t\left(M^T\Gamma_t\right) + \omega_t\right),$$
$$= \mathrm{e}^{\omega_t}P\left(-\omega_t\right), \qquad (11)$$

which is the ESFT. In section 7 we shall use an example to demonstrate the reliance of the ESFT on the velocity symmetry of the initial pdf.

6.4. Symmetry and asymmetry in forward and backward time evolution under a time independent protocol

There is a past–future symmetry in the evolution of $\langle\omega_t\rangle_0$ for positive and negative t if the initial pdf is symmetric. We assume a time independent protocol that satisfies $S_{-t}M^T = M^T S_t$ and write

$$\langle\omega_{-t}\rangle_0 = \int \mathrm{d}\Gamma f\left(\Gamma,\,0\right)\ln\left[\frac{f\left(\Gamma,\,0\right)\mathrm{d}\Gamma}{f\left(S_{-t}\Gamma,\,0\right)\mathrm{d}\left(S_{-t}\Gamma\right)}\right],$$
$$= \int \mathrm{d}\left(M^T\Gamma\right) f\left(M^T\Gamma,\,0\right)\ln\left[\frac{f\left(M^T\Gamma,\,0\right)\mathrm{d}\left(M^T\Gamma\right)}{f\left(S_{-t}M^T\Gamma,\,0\right)\mathrm{d}\left(S_{-t}M^T\Gamma\right)}\right],$$
$$= \int \mathrm{d}\Gamma f\left(\Gamma,\,0\right)\ln\left[\frac{f\left(\Gamma,\,0\right)\mathrm{d}\Gamma}{f\left(M^T S_t\Gamma,\,0\right)\mathrm{d}\left(M^T S_t\Gamma\right)}\right],$$
$$= \int \mathrm{d}\Gamma f\left(\Gamma,\,0\right)\ln\left[\frac{f\left(\Gamma,\,0\right)\mathrm{d}\Gamma}{f\left(S_t\Gamma,\,0\right)\mathrm{d}\left(S_t\Gamma\right)}\right], \qquad (12)$$

so $\langle\omega_{-t}\rangle_0 = \langle\omega_t\rangle_0$. If $f\left(M^T\Gamma,\,0\right) \neq f\left(\Gamma,\,0\right)$ this outcome does not necessarily follow, but nevertheless $\langle\omega_t\rangle_0 = 0$ at $t = 0$ and is non-negative for $t \neq 0$, so there is a local minimum in $\langle\omega_t\rangle_0$ at $t = 0$.

6.5. Covariance

Next we investigate the significance of the reference time at which the initial pdf is defined. We write

$$
\begin{aligned}
\langle \omega_t \rangle_0 &= \int \mathrm{d}\Gamma f(\Gamma, 0) \ln\left[\frac{f(\Gamma, 0)\mathrm{d}\Gamma}{f\left(S_t\Gamma, 0\right)\mathrm{d}\left(S_t\Gamma\right)}\right],\\
&= \int \mathrm{d}\Gamma_\tau f\left(\Gamma_\tau, \tau\right)\ln\left[\frac{f\left(\Gamma_\tau, \tau\right)\mathrm{d}\Gamma_\tau}{f\left(S_\tau S_t\Gamma, \tau\right)\mathrm{d}\left(S_\tau S_t\Gamma\right)}\right],\\
&= \int \mathrm{d}\Gamma_\tau f\left(\Gamma_\tau, \tau\right)\ln\left[\frac{f\left(\Gamma_\tau, \tau\right)\mathrm{d}\Gamma_\tau}{f\left(S_t\Gamma_\tau, \tau\right)\mathrm{d}\left(S_t\Gamma_\tau\right)}\right],\\
&= \langle \omega_t \rangle_\tau.
\end{aligned}
\tag{13}
$$

where $\Gamma_\tau = S_\tau\Gamma$ and we again take the protocol to be time independent. The mean integrated modified dissipation function for a period t is independent of the reference time at which the initial pdf is specified in these circumstances. Similarly, for the unaveraged ω_t we have $\omega_t(\Gamma, 0) = \ln\left[f(\Gamma, 0)\mathrm{d}\Gamma/f(S_t\Gamma, 0)\mathrm{d}(S_t\Gamma)\right] = \ln\left[f(\Gamma_\tau, \tau)\mathrm{d}\Gamma_\tau/f(S_t\Gamma_\tau, \tau)\mathrm{d}(S_t\Gamma_\tau)\right] = \omega_t(\Gamma_\tau, \tau)$, where the second argument of ω_t indicates the reference time.

It is of interest to compare equation (13) with a result referred to as covariance [36]. We take care to show explicitly the M^T operator inside the phase space increment in the denominator (compare with equation (1)). Again adding a second argument to the integrated dissipation function to indicate the reference time, we write

$$
\begin{aligned}
\Omega_{2\tau+t}(\Gamma, 0) &= \ln\left[\frac{f(\Gamma, 0)\mathrm{d}\Gamma}{f\left(M^T S_{2\tau+t}\Gamma, 0\right)\mathrm{d}\left(M^T S_{2\tau+t}\Gamma\right)}\right],\\
&= \ln\left[\frac{f\left(\Gamma_\tau, \tau\right)\mathrm{d}\Gamma_\tau}{f\left(S_\tau M^T S_{2\tau+t}\Gamma, \tau\right)\mathrm{d}\left(S_\tau M^T S_{2\tau+t}\Gamma\right)}\right],\\
&= \ln\left[\frac{f\left(\Gamma_\tau, \tau\right)\mathrm{d}\Gamma_\tau}{f\left(S_\tau M^T S_{\tau+t}\Gamma_\tau, \tau\right)\mathrm{d}\left(S_\tau M^T S_{\tau+t}\Gamma_\tau\right)}\right],\\
&= \ln\left[\frac{f\left(\Gamma_\tau, \tau\right)\mathrm{d}\Gamma_\tau}{f\left(M^T S_t\Gamma_\tau, \tau\right)\mathrm{d}\left(M^T S_t\Gamma_\tau\right)}\right],\\
&= \Omega_t\left(\Gamma_\tau, \tau\right),
\end{aligned}
\tag{14}
$$

and hence $\Omega_0(\Gamma_\tau, \tau) = \Omega_{2\tau}(\Gamma, 0)$ as shown in [36]. The initial value (i.e. when $t = 0$) of the integrated dissipation function referred to time τ is equal to the integrated dissipation function for a period of length 2τ when referred to time zero. This result appears to be a consequence of the special character of the system at zero time where velocity symmetry of the pdf is assumed, as well as the presence of the velocity inversion operator in the definition of Ω_t, but is more complex than equation (13).

Further properties of the modified dissipation function are explored in appendix A.

7. Obversibility test in a simple system

A system characterised by a velocity asymmetric pdf is not an unusual situation and the modified dissipation function can provide a framework for discussing irreversibility in its evolution, by means of the obversibility assessment described in section 4. For example the pdf of a stationary state with non-zero spatial current is likely to be asymmetric in velocity, making the dissipation function inappropriate.

In this section we analyse a simple and familiar model in nonlinear dynamics [37] to illustrate several of the properties of ω_t discussed in section 6. Consider a single particle moving in two dimensions but constrained to have constant kinetic energy. It is acted upon by an external force $\boldsymbol{F}$ that drives the momentum of the particle towards a value $\boldsymbol{p}_0$. It is a deterministic dynamical system satisfying time reversal symmetry, with a time independent protocol, and where all points in phase space (except for $\boldsymbol{p} = -\boldsymbol{p}_0$) flow towards a fixed point at $\boldsymbol{p} = \boldsymbol{p}_0$.

The particle momentum $\boldsymbol{p}$ evolves according to

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F}(\boldsymbol{p}) = \boldsymbol{F}_0 - \frac{\boldsymbol{F}_0 \cdot \boldsymbol{p}}{|\boldsymbol{p}_0|^2}\boldsymbol{p}, \tag{15}$$

such that $\boldsymbol{F}(\boldsymbol{p}) \cdot \boldsymbol{p} = 0$, implying constancy of the energy:

$$\frac{d\,|\boldsymbol{p}|^2}{dt} = 2\boldsymbol{p} \cdot \frac{d\boldsymbol{p}}{dt} = 2\boldsymbol{p} \cdot \boldsymbol{F}(\boldsymbol{p}) = 0. \tag{16}$$

The form of the dynamics has a basis in Gauss' principle of least constraint [16, 17].

We take $\boldsymbol{F}_0$ to be parallel to $\boldsymbol{p}_0$ so that $\boldsymbol{F}(\boldsymbol{p}_0) \times \boldsymbol{p}_0 = 0$, which with $\boldsymbol{F}(\boldsymbol{p}_0) \cdot \boldsymbol{p}_0 = 0$ implies $\boldsymbol{F}(\boldsymbol{p}_0) = 0$ at the fixed point. We multiply equation (16) by $\boldsymbol{p}_0$ and write $\boldsymbol{p}_0 \cdot \boldsymbol{p} = |\boldsymbol{p}|^2 \cos\theta$ so that

$$\begin{aligned} |\boldsymbol{p}|^2\frac{d(\cos\theta)}{dt} &= \boldsymbol{F}_0 \cdot \boldsymbol{p}_0 - \frac{\boldsymbol{F}_0 \cdot \boldsymbol{p}}{|\boldsymbol{p}_0|^2}\boldsymbol{p} \cdot \boldsymbol{p}_0, \\ &= \boldsymbol{F}_0 \cdot \boldsymbol{p}_0 - \boldsymbol{F}_0 \cdot \boldsymbol{p}\cos\theta, \end{aligned} \tag{17}$$

where θ is the angle between $\boldsymbol{p}$ and $\boldsymbol{p}_0$. Then with the transformation $t \to t\,|\boldsymbol{p}\,|^2 / \boldsymbol{F}_0 \cdot \boldsymbol{p}_0$ we have $d(\cos\theta)/dt = 1 - \cos^2\theta$ or $d\theta/dt = -\sin\theta$, for which the solution is

$$\tan[\theta(t)/2] = e^{-t}\tan[\theta(0)/2]. \tag{18}$$

The direction of particle motion turns towards an angle $\theta(\infty) = 0$ as time progresses.

We can determine two equivalent forms for the phase space compression factor:

$$\begin{aligned} \frac{d\theta(t)}{d\theta(0)} &= \frac{\left(1 + \tan^2[\theta(0)/2]\right)e^{-t}}{1 + e^{-2t}\tan^2[\theta(0)/2]} \\ &= \frac{1 + e^{2t}\tan^2[\theta(t)/2]}{\left(1 + \tan^2[\theta(t)/2]\right)e^{t}}. \end{aligned} \tag{19}$$

Furthermore, the probability density function $f(\theta, t)$ satisfies $\partial f/\partial t = \partial(\sin\theta f)/\partial\theta$, and using

$$f[\theta(t), t] = f[\theta(0), 0]\frac{d\theta(0)}{d\theta(t)}, \tag{20}$$

we can solve the dynamics for two interesting initial pdfs at $t = 0$. First, consider case 1 where the initial pdf is uniform, $f_1[\theta(0), 0] = (2\pi)^{-1}$, and hence isotropic and velocity symmetric. We obtain

$$f_1[\theta(t), t] = \frac{\left(1 + \tan^2[\theta(t)/2]\right)e^{t}}{2\pi\left(1 + e^{2t}\tan^2[\theta(t)/2]\right)}. \tag{21}$$

We also consider case 2 where $f_2[\theta(0), 0] = (\pi)^{-1}\cos^2[\theta(0)/2]$, which is anisotropic and velocity asymmetric: we get

$$\begin{aligned} f_2[\theta(t), t] &= \frac{1}{\pi\left(1 + e^{2t}\tan^2[\theta(t)/2]\right)}\frac{\left(1 + \tan^2[\theta(t)/2]\right)e^{t}}{\left(1 + e^{2t}\tan^2[\theta(t)/2]\right)}, \\ &= \frac{\left(1 + \tan^2[\theta(t)/2]\right)e^{t}}{\pi\left(1 + e^{2t}\tan^2[\theta(t)/2]\right)^2}. \end{aligned} \tag{22}$$

As a consequence of the initial isotropy, there is a form of past-future symmetry for case 1, made evident as follows. Define $\phi = \theta + \pi$ such that $\sin\phi/2 = \cos\theta/2$ and $\cos\phi/2 = -\sin\theta/2$. Hence

$$\begin{aligned} f_1(\theta + \pi, -t) &= \frac{\left(1 + \tan^2(\phi/2)\right)e^{-t}}{2\pi\left(1 + e^{-2t}\tan^2(\phi/2)\right)}, \\ &= \frac{\left(1 + \cot^2(\theta/2)\right)e^{t}}{2\pi\left(1 + e^{-2t}\cot^2(\theta/2)\right)e^{2t}}, \\ &= \frac{\left(\tan^2(\theta/2) + 1\right)e^{t}}{2\pi\left(e^{2t}\tan^2(\theta/2) + 1\right)} = f_1(\theta, t). \end{aligned} \tag{23}$$

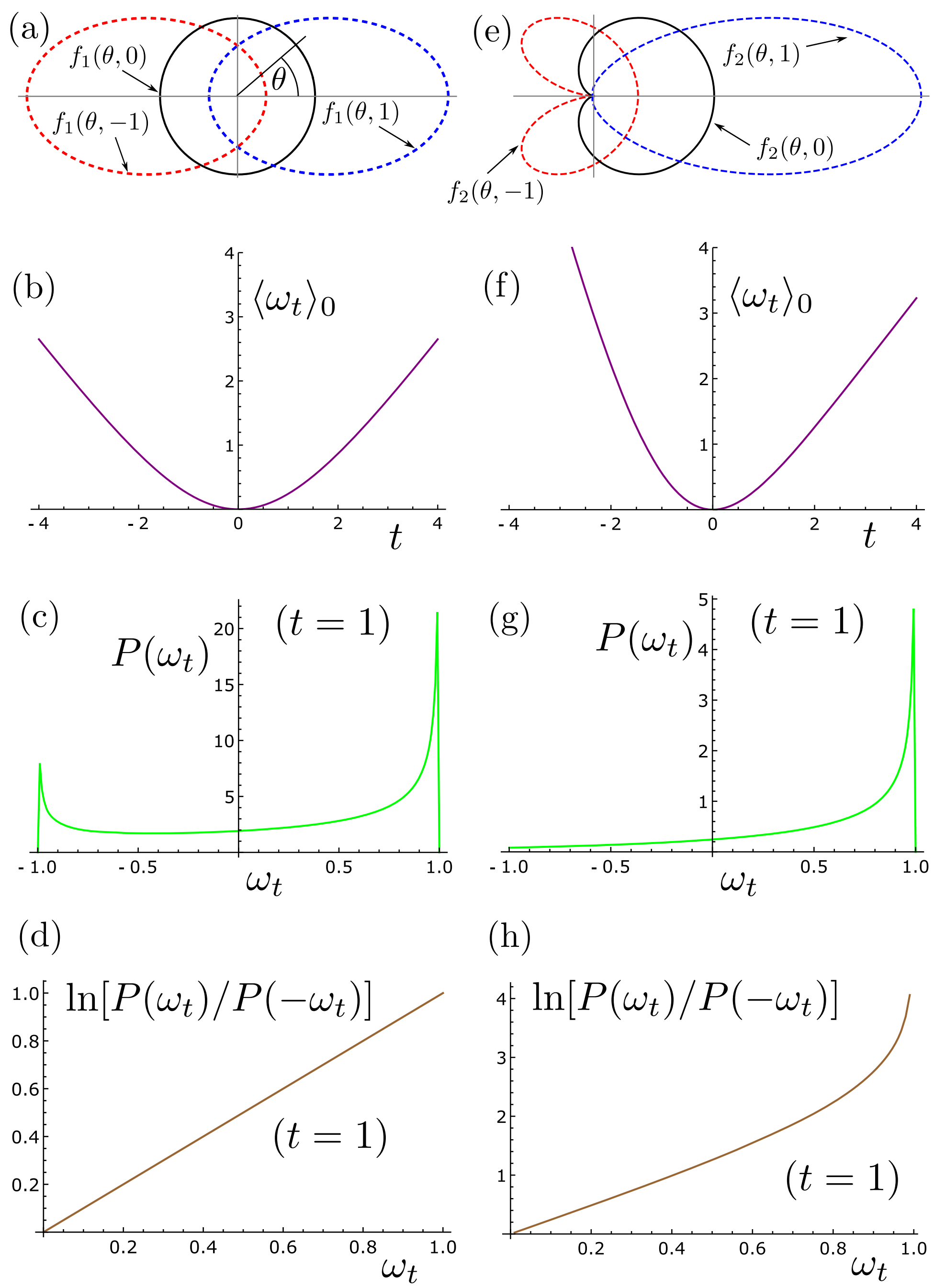

Figure 7. The evolving pdfs as well as the properties of the integrated modified dissipation function are contrasted for case 1 (parts (a)–(d)) and case 2 (parts (e)–(h)), referring to the orientation dynamics of equation (15) with differing initial statistics. At $t = 0$, the pdfs of momentum for case 1 and case 2 are velocity symmetric and asymmetric, respectively.

However, a similar symmetry does not hold for case 2. The pdfs at $t = -1$, 0 and 1 for cases 1 and 2 are illustrated in figures 7(a) and (e), respectively and the presence and absence of symmetry between the past and future of the two situations is clear.

Next we calculate the integrated modified dissipation function for the initially isotropic case 1, defined with respect to a general reference time τ:

$$\omega_t\left[\theta(\tau),\,\tau\right] = \ln\left[\frac{f_1\left[\theta(\tau),\,\tau\right]}{f_1\left[\theta(\tau+t),\,\tau\right]}\frac{\mathrm{d}\theta(\tau)}{\mathrm{d}\theta(\tau+t)}\right]. \tag{24}$$

It is clear from conservation of probability that this is independent of τ and may be written

$$\omega_t[\theta(\tau),\tau] = \omega_t[\theta(0),0] = \ln\left[\frac{f_1[\theta(0),0]}{f_1[\theta(t),0]}\frac{\mathrm{d}\theta(0)}{\mathrm{d}\theta(t)}\right],$$
$$= \ln\left[\frac{\mathrm{d}\theta(0)}{\mathrm{d}\theta(t)}\right] = \ln\left[\frac{\left(1+\mathrm{e}^{-2t}\tan^2[\theta(0)/2]\right)\mathrm{e}^t}{1+\tan^2[\theta(0)/2]}\right], \tag{25}$$

and the mean integrated modified dissipation function is also independent of τ:

$$\langle\omega_t[\theta(\tau),\tau]\rangle_\tau = \int_{-\pi}^{\pi}\mathrm{d}\theta(\tau)f_1[\theta(\tau),\tau]\,\omega_t[\theta(\tau),\tau],$$
$$= \langle\omega_t[\theta(0),0]\rangle_0 = \int_{-\pi}^{\pi}\mathrm{d}\theta(0)f_1[\theta(0),0]\,\omega_t[\theta(0),0],$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta\,\ln\left[\frac{\left(1+\mathrm{e}^{-2t}\tan^2(\theta/2)\right)\mathrm{e}^t}{1+\tan^2(\theta/2)}\right]. \tag{26}$$

The non-negativity of this quantity is illustrated in figure 7(b). For large positive t, we have

$$\langle\omega_t[\theta(0),0]\rangle_0 \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta\,\ln\left(\frac{\mathrm{e}^t}{1+\tan^2(\theta/2)}\right) \approx t, \tag{27}$$

and a similar limiting behaviour emerges for large negative t. Indeed we note that $\langle\omega_{-t}\rangle = \langle\omega_t\rangle$.

We do not always expect the mean integrated modified dissipation function to evolve towards a constant. This would correspond to a stationary state but not all nonlinear dynamical systems embody the action of a thermostat. Our purpose has been instead to demonstrate with a practical example that this quantity can grow symmetrically into both the past and future, and in a fashion irrespective of the starting time. This case is rather analogous to the sequence illustrated in figure 1 with regard to past-future symmetry and the arbitrariness of the reference point in time.

We can construct a pdf of values of the integrated modified dissipation function generated for case 1 after a time t. The pdf is defined by

$$P(\omega_t) = \int_{-\pi}^{\pi}\mathrm{d}\theta f_1[\theta,0]\,\delta\left(\omega_t[\theta,0]-\omega_t\right), \tag{28}$$

with $f_1 = 1/(2\pi)$ and $\omega_t[\theta,0]$ given by equation (25). For $t = 1$, the pdf is shown in figure 7(c), and we plot $\ln[P(\omega_t)/P(-\omega_t)]$ against ω_t in part (d) to demonstrate that it satisfies the ESFT; a consequence of the velocity-symmetric initial pdf and the time independent protocol.

We now turn to case 2 where the initial pdf at $t = 0$ is asymmetric in velocity. The integrated modified dissipation function is

$$\omega_t[\theta(0),0] = \ln\left[\frac{f_2[\theta(0),0]}{f_2[\theta(t),0]}\frac{\mathrm{d}\theta(0)}{\mathrm{d}\theta(t)}\right],$$
$$= \ln\left[\frac{\cos^2[\theta(0)/2]}{\cos^2[\theta(t)/2]}\frac{\left(1+\mathrm{e}^{-2t}\tan^2[\theta(0)/2]\right)\mathrm{e}^t}{1+\tan^2[\theta(0)/2]}\right],$$
$$= 2\ln\left[\frac{\left(1+\mathrm{e}^{-2t}\tan^2[\theta(0)/2]\right)\mathrm{e}^{t/2}}{1+\tan^2[\theta(0)/2]}\right], \tag{29}$$

and the mean integrated modified dissipation function is $\langle\omega_t[\theta(0),0]\rangle_0 = \int_{-\pi}^{\pi}\mathrm{d}\theta(0)f_2[\theta(0),0]\,\omega_t[\theta(0),0]$ with $f_2 = \pi^{-1}\cos^2[\theta(0)/2]$ and $\omega_t[\theta(0),0]$ given by equation (29). This quantity is shown in figure 7(f) to demonstrate its asymmetry with respect to the past and future, a consequence of the velocity asymmetry of the pdf at $t = 0$. The pdf of ω_t for $t = 1$ in figure 7(g) illustrates the distribution of outcomes, and in contrast to case 1, the plot of $\ln[P(\omega_t)/P(-\omega_t)]$ against ω_t is nonlinear (shown as part (h)), demonstrating that the integrated modified dissipation function for this asymmetric initial condition does not satisfy the ESFT.

8. Revisiting the fundamental statistical postulate

Our discussion of asymmetry in the velocity statistics of a system obliges us to revisit the fundamental postulate of statistical physics: the allocation of equal probabilities, or credibility, to all accessible microstates of a system for the purpose of computing equilibrium averages, with the implication that pairs of velocity-inverted configurations are taken to be equally credible.

Now, it is clear that the fundamental postulate is a simplified representation of reality, since allocation of credibility depends on how much conditioning of the statistics is employed. Let us suppose that the present state of a system is conditioned on its having evolved from a situation in the past (the imposition of a Past Hypothesis) as a result of the removal of a constraint. A gas might be released from a small enclosure into a larger container and allowed to come into equilibrium. Liouville's theorem requires that the volume of phase space corresponding to credible configurations in the past situation is preserved for an isolated system evolving according to Hamiltonian dynamics, and so only a tiny proportion of configurations in the wider phase space corresponding to the larger container are actually credible under this conditioning.

For the purpose of determining statistical behaviour, however, this assessment can often be replaced by one where credibility is extended across all available configurations, which is of course the fundamental postulate. It is not quite the same as coarse-graining. It is a (largely) benign act of generosity that makes the averaging easier. The circumstances would require the dynamics to be mixing, such that configurations that are credible according to conditioning on the past are spread uniformly, in some sense, over the wider phase space, and a sufficient period of time will need to have elapsed since the constraint was released. These issues have been considered before (e.g. [13]). Furthermore, the extension of credibility across the phase space would work if the function that is to be averaged is smooth over the phase space. For example, credibility can be extended from one to both members of a velocity-inverted pair of configurations for the purpose of calculating the mean system energy since this property takes the same value in the two configurations. The procedure of extension of credibility is a mathematically convenient make-believe.

Even if a set of probabilities of differing magnitudes had been assigned to the original credible configurations, mixing would still tend to produce a broadly equal density of probability across the wider phase space once equilibrium has been established. The fundamental postulate would still be a very good make-believe representation of the configurational credibility for the purposes of computing equilibrium averages of most functions.

But there are functions whose averages would not be well represented by this extension and smoothing of credibility, and the most obvious is an indicator function that would flag that a configuration had indeed evolved from the specified macrostate in the past. The velocity-inverted partners of configurations that are positively flagged would very likely not have evolved from the original macrostate and would be negatively flagged. The reason is that if a configuration has evolved from a particular macrostate in the past, then its velocity-inverted partner, in the absence of external interactions, will evolve to a velocity-inverted image of that macrostate in the *future*. A configuration that evolves to a physically similar version of the original macrostate will almost surely not have evolved from that macrostate in the past and therefore is highly likely to be negatively flagged. This issue is not the same as a Poincaré recurrence, but nevertheless a return to a (low volume) macrostate in the course of the dynamics should also be ascribed a low probability. We assume that the time elapsed since the release of the constraint is not infinite. The credibility of configurations taking into account conditioning on the past is illustrated in figure 8.

The indicator function is a property of a configuration quite unlike energy. The function is not smooth over the phase space and in particular its value is not likely to be the same for velocity-inverted partners. For the purposes of averaging functions such as this, the fundamental postulate is a very poor representation of the statistical weighting of the configurations. Functions that indicate relaxational behaviour or provide a measure of irreversibility will typically come into this category.

We have considered up to now that an equilibrium system necessarily has evolved from a prior situation and that an appropriate conditioning of the probabilities assigned to each system configuration should be performed. In other words, we have implemented a Past Hypothesis and concluded that velocity asymmetry in the system pdf is to be expected. But a different conditioning can eliminate this expectation and it is worth considering some examples, partly to restore some faith in the fundamental postulate.

For example, it is perfectly reasonable to imagine at some point separating the system into parts, after which they no longer interact with one another. A box containing a gas might be divided into two by the insertion of the traditional partition. There would then be no requirement that a subsystem has differing probabilities of adopting each member of a velocity-inverted pair of configurations because the subsystem essentially has no previous history upon which it can be conditioned. A velocity-inverted version of a given configuration will not evolve forwards in time in a fashion that retraces the previous history of its partner; it cannot do so beyond the point in the past where the system was divided into subsystems. What this reveals is that for a system that is

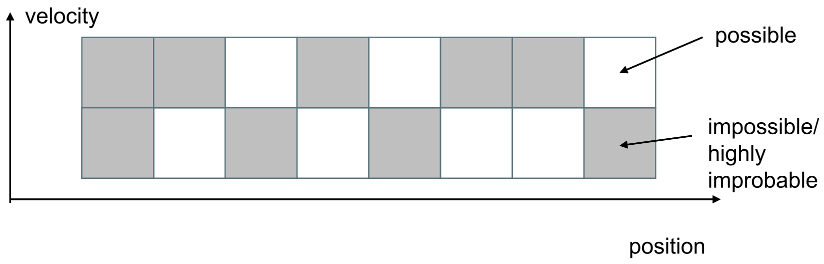

Figure 8. An illustration of the credibility that should be assigned to configurations of a system, conditioned on whether they evolve according to time reversal symmetric dynamics from a particular macrostate in the past (or indeed to one in the future). For each specification of coordinates there is a partner with an inverted set of velocities. By far the most common circumstance is that neither partner evolves from (or to) the specified conditioning macrostate, but where one member of the pair does, the other is highly unlikely to do so as well, since its future is related to its partner's past, and a recurrence of a low volume macrostate along a single trajectory is considered to be very improbable.

subject to external manipulation, such as division into parts, an imposed Past Hypothesis can only refer back to the last time a constraint on the system was changed. This has the conceptual advantage that we can avoid iterating back in time without limit. Eliminating the capacity to invert velocities and make the system retrace its previous history amounts to a breakage of the time reversal symmetry of the dynamics at those moments. The loss of prior conditioning would mean that the fundamental postulate would then be tenable, and standard statistical averaging would follow.

9. Conclusions

Irreversible behaviour remains a puzzle today, over 150 years after it was first quantified using the concept of thermodynamic entropy production. This is certainly the case in the context of classical physics, the focus of our discussions, and remains so from a quantum perspective [38]. The insights offered by Boltzmann [13, 29, 30] based on the evolution of a system through its phase space provide the most appealing framework for a microscopic understanding, while at the same time raising some additional concerns. If irreversibility is simply a consequence of complex but time-reversible dynamics then why does past behaviour not resemble the predictions of development in the future? How is irreversibility to be made distinct from the formation and dissolution of patterns of the kind shown in figure 1? In the simplest presentation of Boltzmann's ideas, it is difficult to avoid the conclusion that entropy is currently at a minimum, quite at odds with reality. This issue is made very plain when we attempt to quantify a measure of irreversibility based on the dynamics and the statistics of the initial state.

There is a certain consensus that an asymmetry in the nature of possible system configurations between the past and future must be the fundamental rationale for irreversible behaviour of the kind alluded to in section 1. This is known generically as the Past Hypothesis [9], and it amounts essentially to a statement that directionality in time is a consequence of very special initial conditions on a grand scale. Collective properties at earlier times, such as an uneven division of energy or space between the components of a system, condition the statistics of configurations at the present time. It is argued that such unevenness is not an expected feature of the world at later times. The resulting velocity asymmetry in the present configurational statistics identifies for us a preferred mean direction of evolution and potentially a measure of irreversibility. We have illustrated conditioning on the past in a lighthearted fashion by considering the dynamics of trains on a rail network, the asymmetry being a limitation to motion in the same (forward) direction for all time once they have been set moving.

These matters place a requirement on the design of a measure of irreversibility for systems governed by time reversal symmetric deterministic dynamics. The measure must work for velocity asymmetric statistics and not just the symmetric situation characteristic of canonical equilibrium. It is eminently possible that different credibilities should be ascribed to velocity-inverted pairs of configurations in a statistical representation of the system. In principle the Past Hypothesis requires this even if the system has relaxed into an equilibrium state, though the asymmetry might be difficult to recognise for very long elapsed times. The intrinsic asymmetry could also be eliminated if the system is prepared in specific ways, for example by separating it from a larger system. If we assume a velocity symmetric probability density function (pdf) we are effectively declaring the past behaviour to be dynamically inaccessible by velocity inversion, or of no interest since equilibrium is long-established.

The mean time-integrated dissipation function, developed by Evans and collaborators, has the appealing property that, like thermodynamic entropy change, its value cannot be negative, though its development in time is not necessarily monotonic [20, 21]. It was designed for use in situations with velocity-symmetric initial statistics, but the definition of the dissipation function can be slightly modified in order to accommodate asymmetric statistics, and we have proposed such a revised form that retains many of the interesting properties.

The mean time-integrated modified dissipation function can be regarded as an indicator of what we have called the *obversibility* of the dynamics and initial statistics, roughly the capacity that a given behaviour and its time-reversed counterpart might be equally likely to develop starting from a chosen moment in time. It is defined in terms of the probabilities that the system might adopt initial configurations from which a trajectory and its time-reversed antitrajectory partner follow under the dynamics, the latter *after* a velocity inversion. With this modification to the original procedure conceived by Evans and his collaborators we can accommodate velocity-asymmetric pdfs. The point is illustrated in figure 5.

The mean time-integrated dissipation function (modified or not) increases into both the past and future, a version of the minimum entropy problem. Evans and collaborators have argued on the basis of causality [16, 39, 40] that the propagation of the dissipation function into the past is physically inadmissible even if it seems mathematically possible.

When we consider stochastic dynamics, however, we do not need to focus such attention on the symmetry or asymmetry of the statistics, nor make an appeal to causal arguments to avoid a minimum entropy problem, since in this framework the past-future asymmetry or Past Hypothesis can be implemented through time reversal asymmetric effective dynamical rules. In this spirit, if the past behaviour were entirely recorded, rather than to be retrodicted, then modelling evolution into the past using stochastic dynamics would be inadmissible and there would be no minimum entropy problem. This is a distinct advantage of working in a framework of stochastic rather than deterministic dynamics.

The concept of mean stochastic entropy production has been developed as a measure of irreversibility in a framework of stochastic dynamics [22], and it emerges from a test of the *reversibility* of the dynamics. Stochastic entropy production is defined in terms of the likelihoods that the system might adopt configurations from which a trajectory and its antitrajectory partner emerge under the dynamics, the latter after a velocity inversion, but in this case the antitrajectory is to be initiated *after* the trajectory has been completed. This distinguishes the tests for obversibility and reversibility, and provides the mathematical difference between the integrated modified dissipation function and the stochastic entropy production. The point is illustrated by comparing figures 5 and 6. Both measures of irreversibility may be used within a framework of stochastic dynamics, but only the test of obversibility is available if the dynamics are deterministic and time-reversal symmetric, since reversibility is then automatic.

For deterministic dynamics, the mean integrated modified dissipation function is non-negative (the essential property of a measure of irreversibility) even if the initial pdf is velocity asymmetric. However, its statistics will satisfy the ESFT under a time symmetric protocol only if the initial pdf is symmetric in velocity, as has been demonstrated using a simple model of isokinetic particle reorientation and figure 7.

The mean stochastic entropy production also cannot become negative, but its evolution is monotonic [41]. Furthermore, it has been shown to map onto the production of thermodynamic entropy in various cases [26]. The stochastic dynamics can accommodate a velocity symmetric pdf while maintaining a fundamental asymmetry between evolution into the past and future, and this measure of irreversibility can satisfy a symmetry (the detailed fluctuation relation) analogous to the ESFT [24].

To conclude, we have discussed the similarities and differences between two measures of irreversibility that have emerged from studies of systems evolving under deterministic and stochastic dynamics. These measures have been developed in order to quantify our understanding of the empirical phenomenon of thermodynamic entropy production and the second law. In particular, we have noted that the test procedures that lie behind each measure are very similar, differing only in the sequence of events considered. Both approaches can accommodate a statistical asymmetry in the dynamical state of the world, the likely underlying basis of the phenomenon of irreversibility, and arguments can be advanced in each case to avoid an apparent minimum entropy problem. We believe that it is important that these connections between the two viewpoints are more widely appreciated.

Acknowledgments

I am grateful to Lamberto Rondoni for hospitality during a visit to Politecnico di Torino when some of these ideas first took shape.

Appendix A. Further properties of the modified dissipation function for asymmetric velocity statistics

We explore whether the modified dissipation function, designed to accommodate a velocity-asymmetric system pdf, satisfies the *dissipation theorem* derived by Evans and collaborators in [42]. Since

$$f\left(\Gamma_t,\, t\right) = f\left(\Gamma_t,\, 0\right)\exp\left(\int_0^t \omega\left(S_s\Gamma\right)\mathrm{d}s\right), \tag{A.1}$$

we can also write

$$\begin{aligned} f\left(\Gamma,\, t\right) &= f\left(\Gamma,\, 0\right)\exp\left(\int_{-t}^0 \omega\left(S_{s+t}\Gamma_{-t}\right)\mathrm{d}s\right),\\ &= f\left(\Gamma,\, 0\right)\exp\left(-\int_0^{-t} \omega\left(S_s\Gamma\right)\mathrm{d}s\right). \end{aligned} \tag{A.2}$$

We consider the average of a function $B(\Gamma)$:

$$\begin{aligned} \langle B(t)\rangle \equiv \langle B(\Gamma)\rangle_t &= \int \mathrm{d}\Gamma f\left(\Gamma,\, t\right)B(\Gamma)\\ &= \int \mathrm{d}\Gamma_t f\left(\Gamma_t,\, t\right)B\left(\Gamma_t\right)\\ &= \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B\left(\Gamma_t\right) = \langle B\left(\Gamma_t\right)\rangle_0. \end{aligned} \tag{A.3}$$

We write

$$\langle B(t)\rangle = \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B(\Gamma)\exp\left(-\int_0^{-t} \omega\left(S_s\Gamma\right)\mathrm{d}s\right), \tag{A.4}$$

and

$$\begin{aligned} \langle B(t + \mathrm{d}t)\rangle - \langle B(t)\rangle &= \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B(\Gamma)\\ &\times\left[\exp\left(\omega\left(S_{-t}\Gamma\right)\mathrm{d}t\right) - 1\right]\exp\left(-\int_0^{-t} \omega\left(S_s\Gamma\right)\mathrm{d}s\right),\\ &\approx \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B(\Gamma)\left[\omega\left(S_{-t}\Gamma\right)\mathrm{d}t\right]\exp\left(-\int_0^{-t} \omega\left(S_s\Gamma\right)\mathrm{d}s\right),\\ &= \int \mathrm{d}\Gamma f\left(\Gamma,\, t\right)B(\Gamma)\left[\omega\left(S_{-t}\Gamma\right)\mathrm{d}t\right],\\ &= \int \mathrm{d}\Gamma_t f\left(\Gamma_t,\, t\right)B\left(\Gamma_t\right)\omega(\Gamma)\mathrm{d}t,\\ &= \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B\left(\Gamma_t\right)\omega(\Gamma)\mathrm{d}t, \end{aligned} \tag{A.5}$$

such that

$$\begin{aligned} \frac{\mathrm{d}\langle B\left(\Gamma_t\right)\rangle_0}{\mathrm{d}t} &= \int \mathrm{d}\Gamma f\left(\Gamma,\, 0\right)B\left(\Gamma_t\right)\omega(\Gamma),\\ &= \langle B\left(\Gamma_t\right)\omega(\Gamma)\rangle_0, \end{aligned} \tag{A.6}$$

which yields the dissipation theorem for the modified dissipation function in the form

$$\langle B\left(\Gamma_t\right)\rangle_0 = \langle B(\Gamma)\rangle_0 + \int_0^t \mathrm{d}s\,\langle B\left(\Gamma_s\right)\omega(\Gamma)\rangle_0, \tag{A.7}$$

analogous to the demonstration by Evans *et al* [42] in terms of Ω.

Weak T-mixing dynamics are defined [19] such that $\lim_{s\to\infty}\langle B(\Gamma_s)\omega(\Gamma)\rangle_0 = \lim_{s\to\infty}\langle B(\Gamma_s)\rangle_0\langle\omega(\Gamma)\rangle_0$. By inserting $B = 1$ into equation (A.6), it can be shown that $\langle\omega(\Gamma)\rangle_0 = 0$, and so for such dynamics a system will clearly relax, acquiring time independent average properties $\langle B(\Gamma_t)\rangle_0$ as $t \to \infty$, an outcome referred to as the *relaxation theorem* [21]. The system discussed in section 7, however, is obviously not an example of T-mixing dynamics.

The assumption of velocity symmetry in the initial system pdf will have the following implication with regard to $\mathrm{d}\langle B(t)\rangle/\mathrm{d}t$. We first investigate the properties of the modified dissipation function under velocity inversion. We assume a time independent protocol such that $S_t M^T = M^T S_{-t}$ for any t. We note from equation (4) that

$$\omega\left(S_t\Gamma\right)\mathrm{d}t = \ln\left[\frac{f\left(S_t\Gamma,\, 0\right)\mathrm{d}\left(S_t\Gamma\right)}{f\left(S_{t+\mathrm{d}t}\Gamma,\, 0\right)\mathrm{d}\left(S_{t+\mathrm{d}t}\Gamma\right)}\right], \tag{A.8}$$

so

$$\omega\left(S_t M^T\Gamma\right)\mathrm{d}t = \ln\left[\frac{f\left(S_t M^T\Gamma,\, 0\right)\mathrm{d}\left(S_t M^T\Gamma\right)}{f\left(S_{t+\mathrm{d}t} M^T\Gamma,\, 0\right)\mathrm{d}\left(S_{t+\mathrm{d}t} M^T\Gamma\right)}\right], \tag{A.9}$$

or

$$\begin{aligned}\omega\left(M^T S_{-t}\Gamma\right)\mathrm{d}t &= \ln\left[\frac{f\left(M^T S_{-t}\Gamma,\, 0\right)\mathrm{d}\left(M^T S_{-t}\Gamma\right)}{f\left(M^T S_{-t-\mathrm{d}t}\Gamma,\, 0\right)\mathrm{d}\left(M^T S_{-t-\mathrm{d}t}\Gamma\right)}\right],\\ &= \ln\left[\frac{f\left(S_{-t}\Gamma,\, 0\right)\mathrm{d}\left(S_{-t}\Gamma\right)}{f\left(S_{-t-\mathrm{d}t}\Gamma,\, 0\right)\mathrm{d}\left(S_{-t-\mathrm{d}t}\Gamma\right)}\right],\\ &= \omega\left(S_{-t}\Gamma\right)(-\mathrm{d}t),\end{aligned} \tag{A.10}$$

implying that $\omega(M^T\Gamma) = -\omega(\Gamma)$ if $f(M^T\Gamma,\, 0) = f(\Gamma,\, 0)$. For a function that is symmetric under velocity inversion, $B(M^T\Gamma) = B(\Gamma)$, this implies that

$$\begin{aligned}\frac{\mathrm{d}\langle B\left(\Gamma_t\right)\rangle_0}{\mathrm{d}t} &= \int \mathrm{d}\Gamma f(\Gamma,\, 0)B\left(\Gamma_t\right)\omega(\Gamma),\\ &= \int \mathrm{d}\left(M^T\Gamma\right)f\left(M^T\Gamma,\, 0\right)B\left(M^T\Gamma_t\right)\omega\left(M^T\Gamma\right),\\ &= -\int \mathrm{d}\Gamma f(\Gamma,\, 0)B\left(\Gamma_t\right)\omega(\Gamma) = 0.\end{aligned} \tag{A.11}$$

This is nothing more than Loschmidt's conclusion that when we consider the evolution of a function (in this case one that is velocity symmetric) under time reversal symmetric dynamics (with no explicit time dependence) starting from an initial state with equal weightings of velocity-inverted configurations, we must infer that any particular change and its opposite are equally likely: the assumptions are consistent only with an equilibrium situation. This is underlined by the further conclusion that the mean of a function that is antisymmetric under velocity inversion is zero in such circumstances, since $\langle B(t)\rangle = \int \mathrm{d}\Gamma f(\Gamma,\, 0)B(\Gamma_t) = \int \mathrm{d}(M^T\Gamma)f(M^T\Gamma,\, 0)B(M^T\Gamma_t) = -\int \mathrm{d}\Gamma f(\Gamma,\, 0)B(\Gamma_t) = 0$. Velocity asymmetry in the initial pdf seems necessary to account for nonequilibrium effects, at least in the absence of a time dependence in the force protocol.

References

[1] Reif F 1965 *Fundamentals of Statistical and Thermal Physics* (New York: McGraw-Hill)
[2] Kondepudi D 2008 *Introduction to Modern Thermodynamics* (Chichester: Wiley)
[3] Ford I J 2013 *Statistical Physics: An Entropic Approach* (Chichester: Wiley)
[4] Orwell G 1945 *Animal Farm* (London: Secker and Warburg)
[5] Coveney P and Highfield R 1991 *The Arrow of Time* (London: Flamingo)
[6] Sklar L 1993 *Physics and Chance* (Cambridge: Cambridge University Press)
[7] Price H 1997 *Time's Arrow and Archimedes' Point* (Oxford: Oxford University Press)
[8] Zeh H D 1999 *The Physical Basis of the Direction of Time* 3rd edn (Berlin: Springer)
[9] Albert D Z 2000 *Time and Chance* (Cambridge, MA: Harvard University Press)
[10] Uffink J 2007 *Handbook of the Philosophy of Physics* ed J Butterfield and J Earman (Amsterdam: North-Holland)
[11] Brown H R, Myrvold W and Uffink J 2009 *Stud. Hist. Phil. Mod. Phys.* **40** 174
[12] North J 2011 *The Oxford Handbook of Philosophy of Time* ed C Callender (Oxford: Oxford University Press)
[13] Hemmo M and Shenker O 2012 *The Road to Maxwell's Demon* (Cambridge: Cambridge University Press)
[14] Earman J 2006 *Stud. Hist. Phil. Mod. Phys.* **37** 399
[15] Evans D J, Cohen E G D and Morriss G P 1993 *Phys. Rev. Lett.* **71** 2401
[16] Evans D J and Searles D J 2002 *Adv. Phys.* **51** 1529
[17] Evans D J and Morriss G P 2008 *Statistical Mechanics of Nonequilibrium Liquids* 2nd edn (Cambridge: Cambridge University Press)
[18] Evans D J, Williams S R and Searles D J 2011 *J. Chem. Phys.* **134** 204113
[19] Evans D J, Williams S R and Rondoni L 2012 *J. Chem. Phys.* **137** 194109
[20] Reid J C, Evans D J and Searles D J 2012 *J. Chem. Phys.* **136** 021101
[21] Reid J C, Williams S R, Searles D J, Rondoni L and Evans D J 2012 *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* ed R J Klages *et al* (Weinheim: Wiley–VCH)
[22] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
[23] Harris R J and Schütz G M 2007 *J. Stat. Mech.* P07020
[24] Seifert U 2008 *Eur. Phys. J.* B **64** 423

[25] Spinney R E and Ford I J 2012 *Phys. Rev.* E **85** 051113
[26] Spinney R E and Ford I J 2012 *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* ed R J Klages *et al* (Weinheim: Wiley–VCH)
[27] Cercignani C 1998 *Ludwig Boltzmann: the Man Who Trusted Atoms* (Oxford: Oxford University Press)
[28] C4 idents http://theident.gallery/c4-2004.php (accessed 11.2.15).
[29] Bricmont J 1996 *The Flight from Science and Reason* ed P Gross *et al* (New York Academy of Sciences) *Ann. New York Acad. Sci.* **775** 131–75
[30] Lebowitz J L 1999 *Rev. Mod. Phys.* **71** 346
[31] Goldstein S 2001 *Chance in Physics: foundations and perspectives* ed J Bricmont, D Dürr, M C Galavott, G Ghirardi, F Petruccione, N Zangh *et al* (Berlin: Springer)
[32] Orban J and Bellemans A 1967 *Phys. Lett.* A **24** 620
[33] Aharony A 1971 *Phys. Lett.* A **37** 45
[34] Schnakenberg J 1976 *Rev. Mod. Phys.* **48** 571
[35] Tomé T and de Oliveira M J 2015 *Phys. Rev.* E **91** 042140
[36] Evans D J, Searles D J and Williams S R 2010 *J. Chem. Phys.* **133** 054507
[37] Rondoni L and Jepps O G 2012 *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* ed R J Klages *et al* (Weinheim: Wiley-VCH)
[38] Horodecki M and Oppenheim J 2013 *Nature Commun.* **4** 2059
[39] Evans D J and Searles D J 1996 *Phys. Rev.* E **53** 5808
[40] Evans D J, Williams S R and Searles D J 2010 *Nonlinear Dynamics of Nanosystems* ed G Radons *et al* (Weinheim: Wiley-VCH)
[41] Carberry D, Reid J C, Wang G, Sevick E, Searles D J and Evans D J 2004 *Phys. Rev. Lett.* **92** 140601
[42] Evans D J, Searles D J and Williams S R 2008 *J. Chem. Phys.* **128** 014504