**Article:**
Jack, RL and Evans, RML (2016) Absence of dissipation in trajectory ensembles biased by currents. Journal of Statistical Mechanics: Theory and Experiment, 2016. 093305. ISSN: 1742-5468

https://doi.org/10.1088/1742-5468/2016/09/093305

# Absence of dissipation in trajectory ensembles biased by currents

**Robert L. Jack**

Department of Physics, University of Bath, Bath BA2 7AY, UK

**R. M. L. Evans**

School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

**Abstract.** We consider biased ensembles of trajectories associated with large deviations of currents in equilibrium systems. The biased ensembles are characterised by non-zero currents and lack the time-reversal symmetry of the equilibrium state. In cases where the equilibrium system has an inversion symmetry which is broken by the bias, we show that the biased ensembles retain a generalised time-reversal symmetry, involving a spatial transformation that inverts the current. This means that these ensembles lack dissipation. Hence, they differ significantly from non-equilibrium steady states where currents are induced by external forces. One consequence of this result is that maximum entropy assumptions (MaxEnt/MaxCal), widely used for modelling thermal systems away from equilibrium, have quite unexpected implications, including apparent superfluid behaviour in a classical model of shear flow.

## 1. Introduction

The mathematical theory of large deviations underlies the rigorous formulation of equilibrium statistical mechanics and thermodynamics [1, 2], showing how the free energy of a very large system is related to the probability of certain rare fluctuations in that system. In addition to the familiar canonical and microcanonical ensembles used in that context, large deviation theory can also be applied to *ensembles of trajectories* [3, 4, 5]. One considers a physical system evolving in time: for long trajectories, ergodicity implies that time-averaged quantities almost always converge to their steady-state averages. Nevertheless, by focussing attention on the rare trajectories in which this convergence does not occur, large deviation theory can reveal unexpected behaviour. Examples include fluctuation theorems [6, 7, 8, 9, 10, 11], and the existence of dynamical phase transitions in both non-equilibrium systems and supercooled liquids [12, 13, 14, 15, 16]. In the context of sheared systems, it has also been proposed that rare trajectories of an equilibrium system can be used to predict its response to shear, beyond the linear-response regime [17, 18, 19, 20].

To study these rare trajectories, it is useful to define new ensembles of trajectories via biases (or constraints) on the dynamical evolution of the original system, so that *typical* trajectories within the new ensembles correspond to the *rare events* of interest in the original system. In this work, we concentrate on the case where the original system is at equilibrium, and the rare trajectories of interest are those where a time-averaged current $J$ has an atypical (non-zero) value. (Here, a current is a generic observable that is odd under time-reversal. Equilibrium states are time-reversal symmetric, so average currents all vanish at equilibrium.)

In cases where the current breaks a spatial inversion symmetry of the equilibrium system, we show that while the biased ensembles support anomalous currents, they do so without dissipation. Motivated by previous work on sheared systems [17, 18, 19, 21], we illustrate our results using a schematic model of a sheared fluid, so the relevant current is the shear rate. However, we frame our main argument in terms of a fairly general Hamiltonian system in contact with a heat bath, and we consider a general class of currents. The presence of the heat bath is not essential when considering biased ensembles, but it is important when comparing such ensembles with systems that are driven by non-conservative external forces.

Briefly, our main result is that for any trajectory that realises a particular current $J$, there is an equally probable trajectory that is obtained by reversing the direction of time, and applying a spatial inversion operation. This result forbids processes such as the flow of heat into the thermal bath: the direction of heat transfer is reversed under time-reversal but is invariant under spatial inversion, so if some trajectory involves heat flowing into the bath, there is an equally probable trajectory where the same quantity of heat flows out of the bath. Hence the average dissipation is zero. (This balance of heat currents is the usual situation at equilibrium, where it follows directly from time-reversal symmetry; here the situation is similar but relies on a symmetry that includes both time-reversal and spatial inversion.) It follows that trajectories of systems sheared by external forces are generically different from the rare large-shear trajectories obtained in biased ensembles. The role of time-reversal symmetry and of currents in this argument means that our results are related to previous work by Maes and co-workers in the context of non-equilibrium response theory [22, 23]. The main consequence of our result is that we identify qualitative differences between responses to external forces on a system, and dynamical biases (or constraints) on time-averaged currents.

The biased ensembles that we consider are also identical [18] to those obtained by Jaynes's maximum entropy inference (MaxEnt) prescription applied to trajectories and using current as a macroscopic observable (in which case it is also known as MaxCal [24, 25, 26]). Hence our results demonstrate that MaxEnt/MaxCal does *not* generally yield the non-equilibrium dynamics of driven systems, contrary to the widely-held hypothesis [24, 25, 26, 18].

## 2. General setting

We consider a system that evolves in time under a dynamics with some stochastic element. We use $x$ to indicate a generic configuration (or phase space point). We concentrate on cases where $x = (\vec{q}, \vec{p})$, with $\vec{q}$ being a vector of generalized co-ordinates and $\vec{p}$ a vector of conjugate momenta. However, the results may be easily generalised to other Markov processes, in which case $x$ might also represent an element of a discrete (or continuous) configuration space.

### *2.1. Equilbrium dynamics*

We first define an equilibrium dynamics and an associated energy function $E(x)$. We fix Boltzmann's constant $k_\mathrm{B} = 1$. We use $X$ to indicate a trajectory of the system, running from an initial time $t = -\tau$ to a final time $t = \tau$. We write $(X)_t = x(t)$ for the state of the system at time $t$. By "an equilibrium dynamics", we mean (i) that the system's dynamical rules have the Boltzman distribution $p_0(x) \propto \mathrm{e}^{-E(x)/T}$ as a steady state, and (ii) that this steady state is time-reversal symmetric. (We further assume that the steady state is unique.) An example is the case where $x = (\vec{q}, \vec{p})$, the energy is

$$E = \sum_i \frac{1}{2} p_i^2 + V(\vec{q}), \tag{1}$$

and the system evoves by Langevin equations

$$\partial_t q_i = p_i\,, \tag{2}$$

$$\partial_t p_i = -\frac{\partial V}{\partial q_i} - \lambda_i p_i + \sqrt{2\lambda_i T}\eta_i\,. \tag{3}$$

Here, $\lambda_i$ is a friction constant and $\vec{\eta}$ a vector of white noises with mean zero and $\langle \eta_i(t)\eta_j(t')\rangle = \delta_{ij}\delta(t-t')$. For $\lambda_i = 0$ we recover Hamiltonian time evolution. We emphasise that the equilibrium steady state associated with this evolution is time-reversal symmetric for all $\lambda_i$, as may be demonstrated explicitly by writing an appropriate Fokker-Planck equation (see Appendix A, below).

It is useful to define an operator $\mathbb{T}$ which gives the time-reversed counterpart of a trajectory $X$. The momenta $p_i$ are odd under time-reversal, so we can write

$$(\mathbb{T}X)_t = \overline{x}(-t), \tag{4}$$

where $\overline{x} = (\vec{q}, -\vec{p})$ is obtained by reversing all momenta in configuration $x$. (The overbar should not be confused with any kind of average.) Considering these dynamics and working in the steady state of the system, one may define a probability density $P_{\rm eq}(X)$ over all possible dynamical trajectories. This distribution has a time-reversal symmetry:

$$P_{\rm eq}(X) = P_{\rm eq}(\mathbb{T}X). \tag{5}$$

### *2.2. Driven dynamics*

Next we define a dynamics where the system is driven out of equilibrium by some external forces. That is, we modify (3) to

$$\partial_t p_i = -\frac{\partial V}{\partial q_i} - \lambda_i p_i + f_i + \sqrt{2\lambda_i T}\eta_i \tag{6}$$

where the external forces $f_i$ may be collected into a vector $\vec{f}$. These forces are assumed to be non-conservative, that is, they cannot be obtained as the gradient of any external potential. The probability distribution for trajectories in the steady state of this non-equilibrium dynamics is denoted by $P_{\rm neq}(X)$. Due to the non-conservative forces, there is no time-reversal symmetry: $P_{\rm neq}(X) \neq P_{\rm neq}(\mathbb{T}X)$.

In this work, we are interested in cases where the external forces $\vec{f}$ break a spatial reflection symmetry of some kind. For example, one might have $E(\vec{q}, \vec{p}) = E(-\vec{q}, -\vec{p})$ so that the system's equilibrium behaviour is unchanged if all co-ordinates are inverted. More generally, define an operator $\mathbb{P}$ by

$$(\mathbb{P}X)_t = \tilde{x}(t), \tag{7}$$

where $\tilde{x}$ is related to $x$ through inversion of one or more co-ordinates (and their conjugate momenta). We assume that the equilibrium dynamics are invariant under this transformation, in which case

$$P_{\rm eq}(X) = P_{\rm eq}(\mathbb{P}X). \tag{8}$$

However, we further assume that the external forces $\vec{f}$ break this symmetry so that $P_{\rm neq}(X) \neq P_{\rm neq}(\mathbb{P}X)$.

Note that the driven dynamics considered here is different from the "driven dynamics" of [17, 18, 19, 20, 21, 27]. We consider here a general non-equilibrium driving force, where they consider a specific force that is chosen so that to mimic particular rare events in the original system.

*2.3. Biased dynamics*

The non-conservative external forces $\vec{f}$ in the driven dynamics will induce currents within the system. We define an instantaneous current $j = j(x)$. The dependence of $j(x)$ on $x$ can be fairly general but in order to be interpreted as a current, we require that it changes sign under time-reversal: $j(\overline{x}) = -j(x)$. A simple case (see below) is that $\vec{f}$ corresponds to a shear stress, in which case the associated current would be a strain rate. The external forces break the spatial reflection symmetry $\mathbb{P}$, and we also assume that $j$ changes sign under this inversion: $j(\tilde{x}) = -j(x)$. The total current associated with a trajectory $X$ is

$$J(X) = \int_{-\tau}^{\tau} j(x(t))\mathrm{d}t. \tag{9}$$

From the symmetry properties of the current, one has $J(\mathbb{T}X) = -J(X)$ and $J(\mathbb{P}X) = -J(X)$.

Following [11, 3, 18], we are concerned here with the large deviations of $J$ in the limit $\tau \to \infty$. To this end we define a biased ensemble of trajectories

$$P_{\mathrm{bias}}(X|\nu) = P_{\mathrm{eq}}(X) \cdot \frac{\mathrm{e}^{\nu J(X)}}{\mathcal{Z}(\nu)} \tag{10}$$

where $\nu$ is the strength of the bias, and $\mathcal{Z}$ is a normalisation constant (or dynamical partition sum). For a physical interpretation of this ensemble, note that for large $\tau$, the ensemble $P_{\mathrm{bias}}$ is very close (in a precise sense [28]) to the ensemble of trajectories obtained by constraining the total current $J$ to some particular value. That is, *typical* trajectories from the biased distribution $P_{\mathrm{bias}}$ are the *least unlikely* trajectories that are consistent with a particular ($\nu$-dependent) value of the total current $J$. Alternatively, (10) is the ensemble with maximum combinatorial entropy relative to $P_{\mathrm{eq}}$, subject to a conditioning on the average current $J$. This is exactly the ensemble that results from Jaynes' MaxEnt or MaxCal procedure [24, 25]. Recently, general properties of ensembles defined as in (10) have been explored in some detail [5, 13, 15, 38, 27]. One important result, already anticipated in [17, 18], is that the ensemble (10) can be generated by a Markov *auxiliary process* [38, 27], but the transition rates for this process involve complicated effective interactions between the components of the system, which cannot be easily calculated or inferred in systems with many degrees of freedom.

Given all these definitions, one easily sees that the probability of a time-reversed trajectory $\mathbb{T}X$ in the biased ensemble is equal to the probability of the original trajectory $X$ in an ensemble with the opposite bias: that is,

$$P_{\mathrm{bias}}(\mathbb{T}X|\nu) = \frac{P_{\mathrm{eq}}(\mathbb{T}X)\mathrm{e}^{\nu J(\mathbb{T}X)}}{\mathcal{Z}(\nu)} = \frac{P_{\mathrm{eq}}(X)\mathrm{e}^{-\nu J(X)}}{\mathcal{Z}(\nu)} = P_{\mathrm{bias}}(X|-\nu). \tag{11}$$

Similarly one finds that $P_{\mathrm{bias}}(\mathbb{P}X|\nu) = P_{\mathrm{bias}}(X|-\nu)$. Hence, substituting $X \to \mathbb{T}X$, one has a "generalised time-reversal" symmetry for the biased ensemble:

$$P_{\mathrm{bias}}(\mathbb{P}\mathbb{T}X|\nu) = P_{\mathrm{bias}}(X|\nu). \tag{12}$$

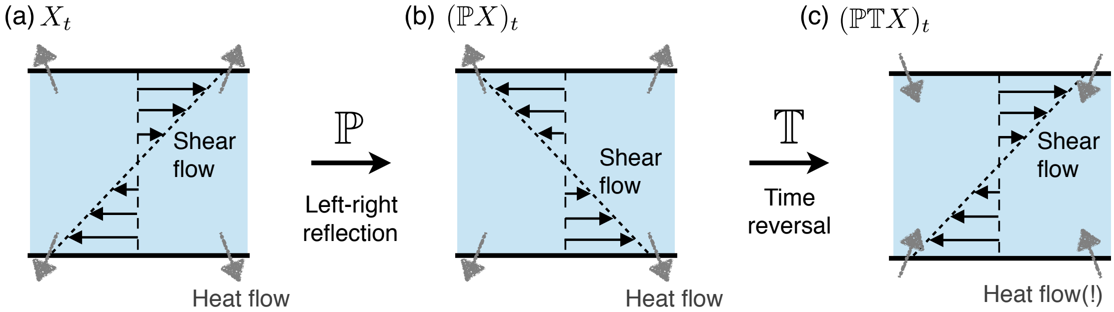


**Figure 1.** (a) Sketch of a sheared system in a dissipative steady state. Black arrows indicate the shear flow of the fluid, while grey arrows indicate heat flow from the fluid into the external heat bath, through the walls of the system. (b) The same system, after left-right reflection, which reverses the shear flow. (c) Reversing the arrow of time reverses both the shear flow and the heat flow, so that energy flows into the system from the bath. For a system that satisfies the $\mathbb{PT}$-symmetry (12), the trajectories illustrated in (a,c) must be equally likely, so the average heat flow must be zero.

That is, given a trajectory $X$, one may obtain another trajectory with equal probability by first inverting the direction of time and then inverting those co-ordinates associated with the operator $\mathbb{P}$. See also [22]. Appendix A illustrates these considerations further, using an operator representation.

Note, we have assumed so far that the system has non-zero frictional and noise forces (i.e., $\lambda_i > 0$), so that its equilibrium steady state is a Boltzmann-distributed state at temperature $T$. However, the analysis leading to (12) holds also for purely deterministic (Hamiltonian) dynamics with $\lambda_i = 0$. The noise and damping forces are useful here since they ensure that the driven system (with $f \neq 0$) eventually converges to a steady state.

### *2.4. Absence of dissipation in the biased ensemble*

Equation (12) is a mathematical statement about trajectory probabilities. To understand its *physical* significance, we now discuss how it is related to heat flow and dissipation. Our central idea is illustrated in Fig. 1, where we sketch the behaviour of a sheared fluid (such examples will be discussed in more detail in Sec. 3). The fluid is confined by two hard walls which we assume to have a well-defined temperature $T$ (these walls might represent, for example, the plates of a rheometer in which an experiment is being performed). Shearing the system leads to dissipation, which appears in the form of heat flow from the fluid into its environment, which we assume acts as a heat bath.

In this system, the symmetry $\mathbb{P}$ corresponds to a plane (left-right) reflection, which inverts the direction of the shear flow, but does not affect the heat flow, as shown in Fig. 1(b). If we then reverse the direction of time, then the shear flow changes direction again, and the direction of the heat flow is also reversed, leading to Fig. 1(c). Equation (12) states that, in the biased ensemble, the situation in Fig. 1(c) must have the same probability as that shown in Fig. 1(a). Hence, on average, there is no heat flow in the biased ensemble.

To derive this result mathematically, we define $\Delta Q(X)$ as the amount of heat that flows from the system into its environment, for trajectory $X$. We emphasise that $\Delta Q(X)$ is a physical quantity which can be measured by calorimetry. From its definition, we have immediately that $\Delta Q(\mathbb{T}X) = -\Delta Q(X)$. This allows us to prove the (trivial) fact that, at equilibrium, there is no heat flow from the system into its environment:

$$\begin{aligned}\langle \Delta Q \rangle_{\rm eq} &= \int {\rm d}X P_{\rm eq}(X) \Delta Q(X) \\ &= \frac{1}{2} \int {\rm d}X P_{\rm eq}(X) [\Delta Q(X) - \Delta Q(\mathbb{T}X)] \\ &= \frac{1}{2} \int {\rm d}X P_{\rm eq}(X) \Delta Q(X) - P_{\rm eq}(\mathbb{T}X) \Delta Q(X) = 0 \qquad (13)\end{aligned}$$

where ${\rm d}X$ indicates an integral over all possible trajectories, the second equality uses $\Delta Q(X) = -\Delta Q(\mathbb{T}X)$; the third equality uses a change of integration varibles $X \to \mathbb{T}X$ and the last equality uses $P_{\rm eq}(X) = P_{\rm eq}(\mathbb{T}X)$.

In both the biased and driven ensembles, the second equality in (13) no longer holds, since $P(\mathbb{T}X) \neq P(X)$. In the driven ensemble, the breaking of this symmetry is typically linked to a finite heat flow $\langle \Delta Q \rangle_{\rm bias}$, which balances (in steady state) the work done on the system by the driving forces. However, in the biased ensemble, the situation is different. On physical grounds (see Fig. 1), we expect the direction of the heat current to be unaffected by the spatial transformation $\mathbb{P}$, so $\Delta Q(\mathbb{P}X) = \Delta Q(X)$. In this case we have

$$\begin{aligned}\langle \Delta Q \rangle_{\rm bias} &= \frac{1}{2} \int {\rm d}X P_{\rm bias}(X) [\Delta Q(X) - \Delta Q(\mathbb{P}\mathbb{T}X)] \\ &= \frac{1}{2} \int {\rm d}X P_{\rm bias}(X) \Delta Q(X) - P_{\rm bias}(\mathbb{P}\mathbb{T}X) \Delta Q(X) = 0 \qquad (14)\end{aligned}$$

where we used the same substitution as in Eq. (13) and the symmetry (12). This is the mathematical statement of the result illustrated in Fig. 1, that the biased ensemble has no average heat flow from the system into its environment. We emphasize that this result does not generally hold for driven systems, as expected on physical grounds since in those systems we do expect heat flow from the system into its environment.

We note in passing that, since the system is coupled to a heat bath at fixed temperature $T$, it seems natural to identify the entropy production for trajectory $X$ as $\Delta S(X) = \beta \Delta Q(X)$ in which case $\langle \Delta S \rangle_{\rm bias} = 0$, again confirming that this ensemble is non-dissipative. However, in contrast to the heat flow $\Delta Q(X)$ which is a physical observable, there is some ambiguity as to the *definition* of the entropy production in the biased ensemble. These issues are discussed in Appendix B, but we emphasise that our conclusion that $\langle \Delta Q \rangle_{\rm bias} = 0$ follows from (12) whenever $\Delta Q(\mathbb{P}X) = \Delta Q(X)$.

### *2.5. Protected observables in the biased ensemble*

As well as the absence of heat flow, the symmetry (12) has several other observable consequences in biased ensembles. In particular, the analysis of the previous section may be generalised to show that there are two classes of observable whose averages must vanish in the biased ensemble, or in any ensemble with the combined $\mathbb{PT}$ symmetry (12). These observables are averages of state-dependent quantities $F(x)$ that are odd under the PT symmetry operation: $F((\mathbb{PT}X)_t) = -F(X_{-t})$.

The first main class includes observables that are odd in $\mathbb{P}$ but even under $\mathbb{T}$. For example, if $\mathbb{P}$ involves inversion of all positions and momenta then any odd function of the position co-ordinates is odd under $\mathbb{PT}$ and must vanish on average in the biased ensemble. That is, if $F = F(q) = -F(-q)$ then $\langle F \rangle_{\rm bias} = 0$, following the same steps as (14).

The second class includes observables that are even in $\mathbb{P}$ but odd under $\mathbb{T}$. For example, if $\mathbb{P}$ involves inversion of just one position co-ordinate $q_1$ and its conjugate momentum $p_1$, then all other momenta $p_2, p_3, \ldots$ are in this class, so that $\langle p_2 \rangle_{\rm bias} = 0$, again by the same argument. This class also includes *dissipative currents* which involve energy flow from the system into a heat bath. Such currents change their sign under $\mathbb{T}$ (that is, reversing the arrow of time means that energy flows from the bath back into the system) but are unchanged by $\mathbb{P}$ (changing the direction of the current $J$ does not affect the direction of the energy flow).

The remaining sections of this paper concentrate on examples inspired by sheared systems, motivated by [17, 18]. Example of observables in the two classes described above are provided in Section 3.4, for one such system.

## 3. Illustrative examples

### *3.1. Linear response*

We first illustrate differences between biased and driven ensembles by considering linear response to the bias $\nu$ and the force $f$. For any observable $O(t)$, we work at equilibrium ($\nu = 0 = f$) and calculate a derivative with respect to $\nu$ (see for example [15]). The result is a fluctuation-dissipation theorem:

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\langle O(0)\rangle_{\rm bias} = \int_{-\infty}^{\infty} C_{\rm eq}^{Oj}(0,t)\mathrm{d}t, \qquad C_{\rm eq}^{Oj}(t',t) = \langle O(t')j(t)\rangle_{\rm eq}, \tag{15}$$

where we use a shorthand notation $j(t) = j(x(t))$, for clarity. Similarly, if we take a force $f$ conjugate to the current $j$, then [29]

$$\frac{\mathrm{d}}{\mathrm{d}f}\langle O(0)\rangle_{\rm neq} = \frac{1}{T}\int_{0}^{\infty} C_{\rm eq}^{Oj}(0,t)\mathrm{d}t. \tag{16}$$

Since the system is at equilibrium, the correlation function $C_{\rm eq}^{Oj}$ depends only on the time difference $t' - t$. The simplest case is $O(t) = j(t)$, in which case we measure the response of the current $j$ itself. In this case the correlation function is also even under time-reversal: $C_{\rm eq}^{jj}(t-t') = C_{\rm eq}^{jj}(t'-t)$. Hence, changing variables $t \to -t$ in (15), one finds $\frac{\mathrm{d}}{\mathrm{d}\nu}\langle j(t=0)\rangle_{\rm bias} = 2T\frac{\mathrm{d}}{\mathrm{d}f}\langle j(t=0)\rangle_{\rm bias}$. A analogous result holds for any observables $O$ for which $C_{\rm eq}^{Oj}(t,t') = C_{\rm eq}^{Oj}(t',t)$, On the other hand, if $C^{Oj}_{\rm eq}(t,t') = -C^{Oj}_{\rm eq}(t',t)$ so that the correlation function is odd under time reversal, then applying again the change of variables $t \to -t$ in (15) yields $\frac{\rm d}{{\rm d}\nu}\langle O(0)\rangle_{\rm bias} = 0$. This result applies if $O$ is a protected observable in the sense of Sec. 2.5, in which case $\langle O(0)\rangle_{\rm bias} = 0$ for all $\nu$. For example, in the first class of protected observable, $O$ depends on position co-ordinates $q$ but not on momenta $p$, so that $O(t) = O(x(t)) = O(\overline{x}(t))$. In that case, time reversal symmetry of the equilbrium state yields $C^{Oj}_{\rm eq}(t',t) = \langle O(t')j(t)\rangle_{\rm eq} = \langle -j(t')O(t)\rangle_{\rm eq} = -C^{Oj}_{\rm eq}(t',t)$, where we used $j(x(t)) = -j(\overline{x}(t))$. Hence $\frac{\rm d}{{\rm d}\nu}\langle O(0)\rangle_{\rm bias} = 0$ for this class of observable. This result is a special case of the general analysis in Sec. 2.5; it shows that the symmetry (12) has observable consequences already in the linear-response regime where the currents flowing in the system are small. (In contrast to this result for the biased ensemble, there is no reason to suppose that the averages of all observables in this class should vanish in the driven ensemble, and the derivatives $\frac{\rm d}{{\rm d}f}\langle O(0)\rangle_{\rm neq}$ are indeed generically non-zero.)

### *3.2. A continuously sheared fluid*

We now illustrate the abstract definitions of the different ensembles by a commonplace example. Fig. 2(a) illustrates a sheared system, for which biased ensembles of the form (10) were discussed in [18, 30]. A slab of fluid sits between two parallel walls, at $y = \pm y_{\rm b}$, with periodic boundaries in the $x$ and $z$ directions. (There should be no confusion between these Cartesian co-ordinates and the notation $x$ for a generic phase space point.) Forces are applied to the plates and the system (eventually) converges to a steady state with a finite shear rate. In this steady state, the external forces are constantly injecting work into the system, this energy acts to heat up the fluid, and eventally flows out through the walls of the system, which we assume to be maintained at constant temperature $T$ by some external thermostat (recall Fig. 1).

The particles within the fluid evolve according to Hamiltonian's equations, except for particles close to the boundary, where they feel (stochastic) thermal noise forces, and shear forces associated with the parallel plates. The equations of motion for the particle momenta can be written in the form (6), except that the thermal noise forces $\eta_i$, damping forces $\lambda_i\omega_i$, and external forces $f_i$ act only at the boundary. In the absence of external forces, one has a time-reversal symmetric steady state.

On introducing a shear stress $\sigma$, a shear rate in the system can be defined as $\dot\gamma = (v_x(y_{\rm b}) - v_x(-y_{\rm b}))/2y_{\rm b}$ where $v_x(y)$ is the average of the $x$-component of the velocity of the fluid, within a thin slab at height $y$. This shear rate will correspond to the general 'current observable' of the previous section: $j = \dot\gamma$. It is a linear combination of velocities, so is manifestly odd under time-reversal, as required. The total shear $\gamma$ is then easily obtained by a time integral, and we identify the time-integrated current $J = \gamma = \int_{-\tau}^{\tau} j(t){\rm d}t$, as in (9). To apply our general discussion to this system, the relevant spatial inversion symmetry $\mathbb{P}$ is the co-ordinate transform $(x, y, z) \to (-x, y, z)$. The equilibrium dynamics are invariant under this transformation; operation with $\mathbb{P}$ inverts the velocities $v_x$ so it also takes $j \to -j$ as required.

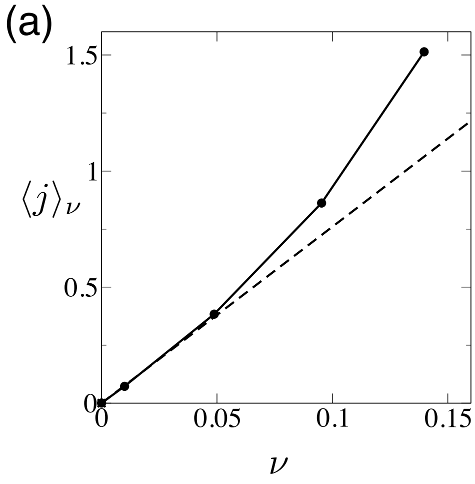


**Figure 2.** (a) Schematic of a sheared system between two parallel plates at $y = \pm y_{\rm b}$, with forces $F$ and $-F$ applied to the top and bottom rates. The mean velocity at height $y$ is $v_x(y)$ with $v_x(y) = y\dot{\gamma}$ in a state of uniform shear rate. We imagine periodic boundaries in the $x$ and $z$ directions. (b) Simplified 'rotor' model, consisting of a set of discs placed along the $y$-axis. The angular velocity of the disc at position $y$ is $\omega_y$, which is analogous to the velocity $v_x(y)$ in (a).

It follows that the generalised time reversal symmetry (12) holds for this system. That is, the biased ensemble of trajectories (10) for this model is invariant under time-reversal followed by a spatial reflection in the plane $x = 0$.

To see the connection of this result to dissipation, we compare this ensemble with a driven (sheared) steady state. In the driven system, one expects currents of energy to flow through the system: the work done by external forces injects energy into the system, this energy flows into the microscopic degrees of freedom of the fluid, and eventually leaves the system as heat, via the external boundaries. If one could reverse the arrow of time, these dissipative energy currents would be reversed: heat would flow into the fluid at the boundaries and appear to perform work on the external plates. A subsequent spatial reflection through $x = 0$ does not reverse the direction of these energy currents. Thus, the driven steady state (with finite shear stress) does not respect the symmetry (12).

If follows that the dissipative energy currents that naturally flow in driven systems are inconsistent with the symmetry relation Eq. (12), so they are forbidden within the biased ensemble (10). This is the sense in which biased ensembles such as (10) differ from driven non-equilibrium ensembles in which external forces act at the boundaries.

### *3.3. A model sheared system*

To make these arguments concrete, we analyse a simple model system in which Eq. (12) has important consequences. We consider a set of $N$ rotors (similar to that in [20]), each with moment of inertia $I$, as illustrated in Fig. 2(b). We draw an analogy between the rotor velocity $\omega_y$ and the velocity $v_x(y)$ for the sheared system shown schematically in Fig. 2. This one-dimensional set of rotors can then be regarded as a highly simplified model of the interactions within a sheared fluid.

In analogy with the interparticle forces in a classical fluid, rotors apply purely conservative torques $u'(\Delta\theta) = \varepsilon \sin(\Delta\theta)$ to their neighbours, that depend only on the relative angle $\Delta\theta_i \equiv \theta_{i+1} - \theta_i$. To model the application of shear stress and heat on the boundaries of the fluid, we apply an additional external torque $f_{\rm t}(t)$ to the topmost rotor, and $f_{\rm b}(t)$ to the rotor at the bottom. The equations of motion are

$$\begin{aligned} I\partial_t\omega_1 &= u'(\Delta\theta_1) + f_{\rm b}(t) \\ I\partial_t\omega_2 &= u'(\Delta\theta_2) - u'(\Delta\theta_1) \\ &\dots \\ I\partial_t\omega_i &= u'(\Delta\theta_i) - u'(\Delta\theta_{i-1}) \\ &\dots \\ I\partial_t\omega_N &= -u'(\Delta\theta_{N-1}) + f_{\rm t}(t). \end{aligned} \tag{17}$$

These equations fully specify the properties of the rotors. The boundary forces $f_{\rm t,b}$ follow from properties of the thermal bath to which the rotors are coupled. They have both deterministic and stochastic parts, arising from applied macroscopic shear stress and heat exchange respectively. We write

$$\begin{aligned} f_{\rm t} &= \lambda_0\Omega - \lambda_0\omega_N + \eta_{\rm t}(t)\sqrt{2\lambda_0 T}, \\ f_{\rm b} &= -\lambda_0\Omega - \lambda_0\omega_1 + \eta_{\rm b}(t)\sqrt{2\lambda_0 T}, \end{aligned}$$

where $\lambda_0$ is a friction coefficient associated with the dissipative coupling to the boundary, $\lambda_0\Omega$ is the external torque on the system, and $\eta_{\rm t}(t)$ and $\eta_{\rm b}(t)$ are independent random noises, with coefficients chosen to respect the Einstein relation for a heat bath of temperature $T$.

At equilibrium ($\Omega = 0$), the $\eta_{\rm t,b}$ are the usual Gaussian noises, and the system is time-reversal symmetric. In the driven case, $\Omega$ is non-zero, while the noises have the same form as at equilibrium. In that case, work is done on the system by the applied torques at the boundaries, which leads to average shear flow. At the same time, heat energy (in the form of disordered motion) flows to the boundaries where it is dissipated. The system will converge to a steady state in which these energy fluxes balance.

In the biased case, no explicit driving force is applied, so $\Omega = 0$, but Eq. 10 means that the noise from the heat bath is sampled non-uniformly, so that the stochastic functions $\eta_{\rm t,b}(t)$ can acquire non-zero expectation values, which induce shear flow. On the face of it, one might imagine that $\langle\eta_{\rm t}\rangle$ in the biased ensemble plays the same role as $\Omega\sqrt{\lambda_0/2T}$ in the driven ensemble, in which case the biased and driven ensembles would be similar. In fact, the two ensembles behave very differently, as we shall now see.

Whatever the ensemble, the mean (time-averaged) torque applied at the top boundary is $\langle f_{\rm t}\rangle = \langle I\partial_t\omega_N + u'(\Delta\theta_{N-1})\rangle$ and, since $\langle\partial_t\omega_N\rangle = 0$ in a steady state, we have

$$\langle f_{\rm t}\rangle = \varepsilon\langle\sin(\Delta\theta_{N-1})\rangle \tag{18}$$

for any steady-state ensemble. (There is also a similar expression for $\langle f_{\rm b}\rangle$.) This means that the mean torque on the boundary can be obtained from the ($i$-dependent) distribution $P(\Delta\theta_i)$ of relative angles between neighbouring rotors. To make progress, we define the symmetry operation $\mathbb{P}$ as the co-ordinate transformation $(\theta_i) \to (-\theta_i)$, which has the properties specified in Sec. 2.2. Also note that reversing the arrow of time leaves $P(\Delta\theta_i)$ unchanged, while the symmetry operation $\mathbb{P}$ changes the sign of $\Delta\theta$. Hence the combined $\mathbb{PT}$ operation transforms $P(\Delta\theta_i)$ to $P(-\Delta\theta_i)$. From (12), the biased ensemble is invariant under $\mathbb{PT}$ so $P(\Delta\theta_i) = P(-\Delta\theta_i)$ within this ensemble (for all $i$). That is, the distribution of $\Delta\theta_i$ is symmetric in the biased ensemble, so $\langle \sin\Delta\theta_i\rangle_{\rm bias} = 0$. Indeed, this last result already follows from the analysis of Sec. 2.5, since $\sin\Delta\theta_i$ is odd under $\mathbb{P}$ but depends only on position co-ordinates, and so is a member of the first class of protected observables discussed in Sec. 2.5.

From here, the startling implication of (18) is that the mean applied torque on the boundary must vanish, $\langle f_{\rm t}\rangle = 0$ in the biased ensemble, thus describing a thermodynamic system induced to flow (shear) continuously by the application of no mean force at all. The system, in the biased ensemble, thus behaves like a superfluid, which is not consistent with the responses of classical systems to external driving.

### *3.4. A sheared model with internal noise*

Our final example is a modified version of the above model, similar to those considered in [31, 21]. In contrast to the previous section, all the rotors are coupled to the thermal bath. For the purposes of this work it is sufficient to consider a system with just three rotors – this very simple system is already sufficient to illustrate the symmetry (12) of biased ensembles, and the breaking of this symmetry in driven systems. It is also simple enough that numerical results are easy to obtain.

As before, the co-ordinates of the system are the angles $\theta_1, \theta_2, \theta_3$ which specify the orientation of the rotors. Each rotor has moment of inertia $I$ so the momenta in the system are $I\omega_i$ with $\omega = \dot{\theta}_i$. The energy of the system is

$$E = \sum_i \frac{1}{2} I\omega_i^2 + u(\theta_1 - \theta_2) + u(\theta_2 - \theta_3) \tag{19}$$

with $u(\Delta\theta) = -\varepsilon\cos\Delta\theta$. Frictional forces act on the velocity differences between all rotors, and a constant driving torque of strength $\sigma$ is applied to the boundary rotors, so that the equations of motion are

$$\begin{aligned} I\partial_t\omega_1 &= -u'(\theta_1 - \theta_2) - \lambda(\omega_1 - \omega_2) + \sqrt{2\lambda T}\eta_1 - \sigma \\ I\partial_t\omega_2 &= -u'(\theta_2 - \theta_3) - u'(\theta_2 - \theta_1) - \lambda(2\omega_2 - \omega_1 - \omega_3) + \sqrt{2\lambda T}(\eta_2 - \eta_1) \\ I\partial_t\omega_3 &= -u'(\theta_3 - \theta_2) - \lambda(\omega_3 - \omega_2) - \sqrt{2\lambda T}\eta_2 + \sigma \end{aligned} \tag{20}$$

where $u'(\Delta\theta) = \varepsilon\sin\Delta\theta$ is the derivative of $u$, and $\eta_{1,2}$ are uncorrelated Gaussian noises with mean zero and variances $\langle\eta_i(t)\eta_j(t')\rangle = \delta_{ij}\delta(t-t')$, as above. Clearly $\partial_t(\omega_1 + \omega_2 + \omega_3) = 0$ so we fix the global momentum to zero without loss of generality: $\omega_1 + \omega_2 + \omega_3 = 0$.

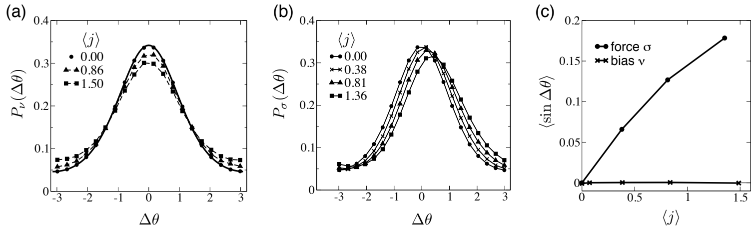


**Figure 3.** Dependence of the normalised shear rate $\langle j \rangle = \langle \omega_3 - \omega_1 \rangle/2$ on bias $\nu$ (a) and applied torque $\sigma$ (b). The unit of time is $\tau_0 = 1$. The dashed lines are linear response results, obtained by numerical evaluation of the correlation functions in (15,16). Expanding about the equilibrium state, one has $\frac{\mathrm{d}}{\mathrm{d}\nu}\langle j(0) \rangle_{\mathrm{bias}} = 2T\frac{\mathrm{d}}{\mathrm{d}\sigma}\langle O(0) \rangle_{\mathrm{neq}}$.

For $\sigma = 0$ one has an equilibrium state with time-reversal symmetry. The system is also invariant under inversion of all positions and momenta: that is, the symmetry operation $\mathbb{P}$ is defined by taking $\tilde{x} = (-\vec{\theta}, -\vec{\omega})$. The shear rate is $j = (\omega_3 - \omega_1)/2$ which is odd under both time-reversal and under $\mathbb{P}$. (The factor of 2 comes from the linear extent of the system along the $y$-direction, for a system of $N$ rotors one would have $j = (\omega_N - \omega_1)/(N-1)$.) Thus, defining a biased ensemble according to (10) with $J = \frac{1}{2}\int_{-\tau}^{\tau}(\omega_3 - \omega_1)\mathrm{d}t$, the generalised time-reversal symmetry (12) applies in this system.

The behaviour of the model is controlled by three dimensionless parameters. The first two of these are $\varepsilon/T$ and $\sigma/T$, which set the strength of the conservative forces and the external forces, respectively. The final parameter is $\lambda_0 = \lambda/\sqrt{IT}$ which sets the strength of the damping. The rotor co-ordinates $\theta$ are naturally dimensionless so it remains only to fix a time unit. There are several intrinsic time scales within the system: we focus on $\tau_0 = I/\lambda$, which is equal to the velocity relaxation time in the weak-force limit $\varepsilon/T \to 0$. When showing numerical results we use units such that $\tau_0 = 1$. This time scale is natural for systems with intermediate damping strength and moderate values of $\varepsilon/T$. Other time scales are more relevant for very strong damping ($\tau_{\mathrm{B}} = \lambda/T = \tau_0\lambda_0^2$); for very weak damping ($\tau_{\mathrm{th}} = \sqrt{I/T} = \tau_0\lambda_0$); or very strong conservative forces ($\tau_{\mathrm{harm}} = \sqrt{I/\varepsilon} = \tau_0\lambda_0\sqrt{T/\varepsilon}$).

*3.4.1. Structure in sheared states* We analyse this model using numerical simulation, in two cases: (i) a non-equilibrium ensemble which depends on the driving force $\sigma$; and (ii) the biased ensemble (10) which depends on the bias strength $\nu$. We consider only the case where $\varepsilon/T = 1$ and $\lambda_0 = 0.3$, which is a representative state point that is sufficient to illustrate our main results. For equilibrium simulations and for case (i), we use solve the equations of motion by the method of Bussi and Parrinello [32], as described in [33]. The time step is fixed at $0.01\tau_0$. For biased ensembles, we use the same scheme in conjunction with transition path sampling methods [34], which are natural tools for sampling ensembles of the form of (10), see for example [16, 35, 36]. We consider trajectories of length $2\tau$ with $\tau = 15\tau_0$, which provides a balance between convergence of the large-$\tau$ limit (as required for studies of large deviations), and manageable computational cost.

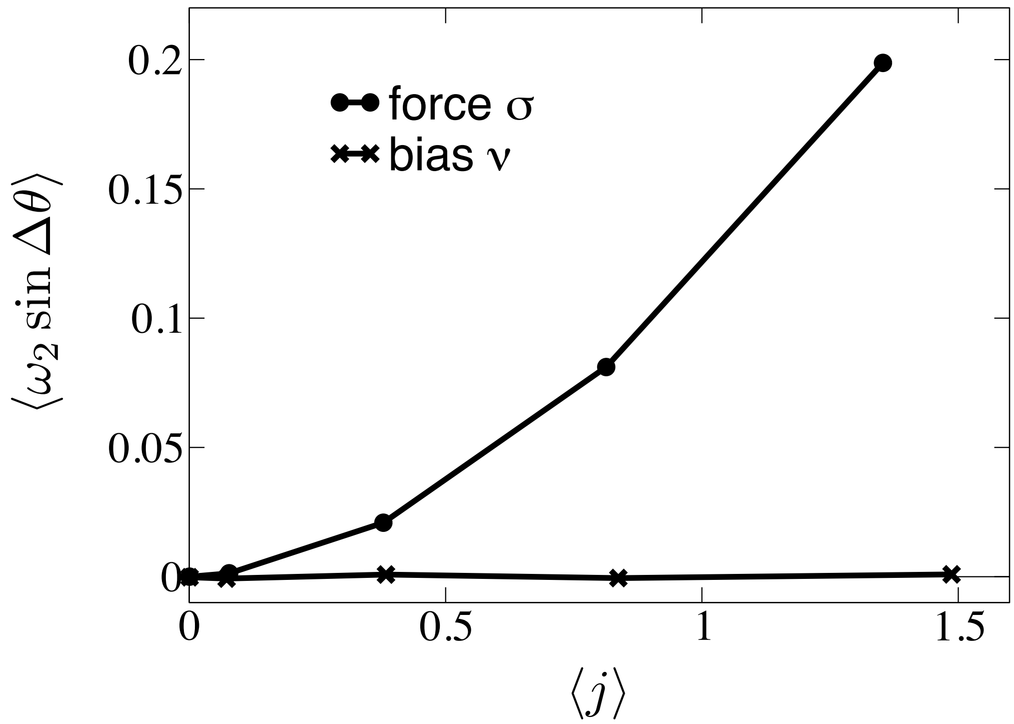


**Figure 4.** (a) Distributions of the angular difference $\Delta\theta = (\theta_2 - \theta_1) \bmod 2\pi$ in biased ensembles with $0 \leq \langle j \rangle < 1.5$, as labelled. The solid line is $P(\Delta\theta) \propto \mathrm{e}^{(\varepsilon/T)\cos\theta}$, dashed lines are guides to the eye. (b) Distributions of $\Delta\theta$ for driven ensembles ($\sigma > 0$), over a similar range of $\langle j \rangle$ (all lines are guides to the eye). As discussed in the main text, the distribution in the biased ensemble is symmetric while the distribution in the driven ensemble is not. (c) Mean conservative force $\langle \sin \Delta\theta \rangle$ plotted parametrically as a function of the current, in both biased and driven ensembles. In the biased case, the symmetry of $P(\Delta\theta)$ means that the average force is always zero.

Note also that the symmetry relation (12) applies for all $\tau$, not only in the large-$\tau$ limit. However, the biased ensemble can be identified with a steady state only when $\tau$ is large [15, 27].

Fig. 3 shows how the shear rate $\langle j \rangle$ depends on the applied bias $\nu$ and applied force $\sigma$. To investigate the structure of the system at finite shear rate, we measure the distribution of the angular difference $\Delta\theta = (\theta_2 - \theta_1)\,\mathrm{modulo}\,2\pi$. At equilibrium $P(\Delta\theta) \propto \mathrm{e}^{\varepsilon \cos\Delta\theta/T}$, consistent with the Boltzmann distribution.

Fig. 4 shows corresponding distributions for the biased and driven ensembles, over comparable ranges of the shear rate $\langle j \rangle$. The distributions differ qualitatively: for the biased state, $P(\Delta\theta)$ is a symmetric function of $\Delta\theta$ while for the driven state, this symmetry is lacking. To further accentuate this difference, we consider the mean force between the rotors $\varepsilon\langle \sin(\theta_2 - \theta_1)\rangle$. For a direct comparison, we plot the mean force parametrically against the shear rate $\langle j \rangle$. The force is a protected observable of the first class described in Sec. 2.5, since it depends only on co-ordinates that are even under time-reversal, but is odd under the parity transformation $\theta_i \to -\theta_i$. For this reason, the average force vanishes in the biased ensemble (consistent with the symmetry of $P(\Delta\theta)$), but is finite in the driven ensemble: hence that driven ensemble does not have the PT-symmetry (12). The symmetry of $P(\Delta\theta)$ in the biased ensemble is also responsible for the vanishing of the mean torque discussed in section 3.2. Since the symmetry of $P(\Delta\theta)$ in the biased case follows from (12), the numerical results in Fig. 4 illustrate the effect of this generalised time-reversal symmetry. The driven system ($\sigma > 0$) lacks the symmetry (12), as is clear from the asymmetry of $P(\Delta\theta)$ in Fig. 4(b).

*3.4.2. Relation to dissipation* To illustrate the relation of these results to dissipation, we consider an observable in the second class discussed in Sec. 2.5, which is a dissipative current. The conservative part of the torque applied to the second rotor by the first is $-u'(\theta_2 - \theta_1)$, so the first rotor does work on the second at a rate $\dot{W}_{12} = -\omega_2 u'(\theta_2 - \theta_1)$. We can interpret this as a flow of energy from rotor 1 to rotor 2. It is manifestly odd under time reversal (since $\omega_2$ is odd) but even under the parity symmetry (since both the angular velocity and the force change sign under that operation). Hence, the PT-symmetry (12) means that $\langle \dot{W}_{12} \rangle_{\text{bias}} = 0$. However, for the driven ensemble, we have generically $\langle \dot{W}_{12} \rangle_\sigma > 0$. Note this quantity is positive, independent of the sign of $\sigma$: the sign of the dissipation is independent of the direction of the applied force, as expected.

Fig 5 shows numerical results for $\dot{W}_{12}$, plotted parametrically as a function of the shear rate. As expected from the discussion in Sec. 2.5, there is no dissipative current in the biased ensemble. In the driven ensemble, the symmetry (12) is broken. It is useful to be clear about the the flow of energy in this case: the external forces do work on the outer rotors $\theta_1, \theta_3$, at rate $\dot{W}_1 = \sigma\omega_1$ and $W_3 = -\sigma\omega_3$. The outer rotors do work on the inner rotor at rates $\dot{W}_{12}$ and $\dot{W}_{32}$. Eventually, all the work done by the outer rotors flows out into the heat bath, through the frictional coupling terms (proportional to $\lambda$). All these heat flows are odd under $\mathbb{T}$ but even under $\mathbb{P}$ so they cannot lead to any average energy transfer in the biased ensemble, although they are all finite in the driven ensemble. [We note in passing that since the system is in a steady state, we have $\partial_t \langle \cos(\theta_2 - \theta_1) \rangle = 0$ even for $\sigma > 0$, and hence $\langle (\omega_1 - \omega_2) \sin(\theta_2 - \theta_1) \rangle = 0$. Hence one always has $\langle \omega_1 \sin(\theta_2 - \theta_1) \rangle = \langle \omega_2 \sin(\theta_2 - \theta_1) \rangle$, the question is whether these two quantities vanish individually, or not.]

*3.4.3. Force balance and non-zero stochastic forces* Finally, it is instructive to take the average of Eq. 20 in the biased ensemble, to make contact with Sec. 3.3. For the first rotor we obtain

$$0 = \varepsilon \langle \sin(\theta_2 - \theta_1) \rangle_{\text{bias}} + \lambda \langle \omega_2 - \omega_1 \rangle_{\text{bias}} + \sqrt{2\lambda T} \langle \eta_1 \rangle_{\text{bias}} \tag{21}$$

For $\nu > 0$ then clearly $\langle \omega_2 - \omega_1 \rangle_{\text{bias}} > 0$, but as noted above, $\langle \sin(\theta_2 - \theta_1) \rangle_{\text{bias}} = 0$. The sum of the last two terms on the right hand side of (21) is analogous to the average force $\langle f_{\text{t}} \rangle_{\text{bias}}$ in Sec. 3.3, and this average force is zero, as noted in that section. Since $\langle \omega_2 - \omega_1 \rangle_{\text{bias}} > 0$, it must therefore be that the noise term has a non-zero average within the biased ensemble

$$\langle \eta_1 \rangle_{\text{bias}} = -\sqrt{\frac{\lambda}{2T}} \langle \omega_2 - \omega_1 \rangle_{\text{bias}}. \tag{22}$$

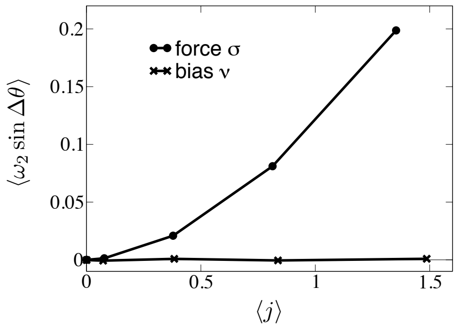


**Figure 5.** Average energy current $\dot{W}_{12} = \langle \omega_2 \sin(\theta_2 - \theta_1) \rangle$, comparing biased ($\nu > 0$) and driven ($\sigma > 0$) ensembles. The current is plotted parametrically against the shear rate $\langle j \rangle$. The numerical results are consistent with the absence of dissipation in the biased ensemble. In the driven ensemble, note that this current is unchanged by the spatial transform $\mathbb{P}$ so it is an even function of $\sigma$, and $d\dot{W}_{12}/d\langle j \rangle = 0$ at $\langle j \rangle = 0$. This contrasts with the mean force shown in Fig. 4c, which changes its sign under $\mathbb{P}$, and is an odd function of $\langle j \rangle$.

Thus, as noted in Sec. 3.3, the finite shear rate that appears in the biased ensemble is sustained by a finite value for a thermal noise force, due to the presence of the bias.

## 4. Conclusion

The main result of this work is Eq. (12), which is a $\mathbb{PT}$-symmetry of biased ensembles of trajectories. Our discussion shows that this symmetry places significant constraints on the behaviour that can be observed in these ensembles. In particular, there is a class of *protected observables* whose average value is always zero, even when currents are flowing in the system. These protected observables are related to dissipative processes in the system, and we argue that their absence means that biased ensembles are non-dissipative.

Ensembles of trajectories of the form (10) appear naturally in calculations based on maximum-entropy inference, since they provide the most likely (or least unlikely) trajectories that are consistent with constraints that are applied to time-integrated currents [18]. The popular MaxCal procedure aims to model non-equilibrium driven systems by using these ensembles of trajectories. Our results show that if we condition on a current that breaks an inversion symmetry, then MaxCal can be valid only if the driven system is $\mathbb{PT}$-symmetric. Given the physical picture illustrated in Fig. 1, we argue that typical driven systems do not have this symmetry, which renders MaxCal invalid in those cases. However, there may be special cases where this symmetry still holds in driven systems, as discussed in Appendix B. In these special cases, our results do not invalidate MaxCal.

From a physical perspective, it is not clear to us *why* these biased ensembles should be free from dissipation. This is a consequence of the time-reversal symmetry of the equilibrium state that survives even in these far-from-equilibrium biased ensembles; it is conditional on the existence of an inversion symmetry of the model, which is broken by the bias (and the drive). We hope that further work on the properties of large deviations in non-equilibrium systems might lead to insights in this direction. For example, the absence of dissipation is related to the response theory of [23] and might also be connected to the effective interactions that arise in biased ensembles [41].

## Acknowledgments

We thank Peter Sollich for helpful discussions. RLJ was supported by the EPSRC through grant EP/I003797/1.

## Appendix A. Operator representations of the generalised time-reversal symmetry

As discussed in [11, 12, 13], biased ensembles of the form (10) are related to "tilted" generators or master operators. The symmetry (12) has a simple interpretation in terms of these operators. We give a brief discussion of this interpretation here (an alternative approach based on path integrals and action functionals can also be used to obtain similar results [22, 23]).

Our starting point is the master operator (the adjoint of the generator) of the equilibrium stochastic process of interest. To analyse the case given in (3), we introduce a representation of the phase space of the system based on Dirac kets $|x\rangle$. The probability distribution $P(x)$ for system's phase space point corresponds to a ket $|P\rangle = \int \mathrm{d}x P(x)|x\rangle$ which evolves in time according to $\partial_t|P\rangle = \mathbb{W}_{\mathrm{eq}}|P\rangle$ with [37]

$$\mathbb{W}_{\mathrm{eq}} = \sum_i \left[ -p_i \frac{\partial}{\partial q_i} + \left( \frac{\partial E}{\partial q_i} + \lambda p_i \right) \frac{\partial}{\partial p_i} + \lambda \left( 1 + T \frac{\partial^2}{\partial p_i^2} \right) \right] \tag{A.1}$$

Applying this operator to the equilibrium (Boltzmann) distribution yields $\mathbb{W}_{\mathrm{eq}}|P_{\mathrm{eq}}\rangle = 0$, confirming that this is indeed the steady state of the model. To analyse the time-reversal symmetry of this model, we introduce an operator $\hat{\mathbb{T}}$ which inverts the direction of momenta: $\hat{\mathbb{T}}|x\rangle = |\overline{x}\rangle$. We introduce a second operator $\hat{\pi}$ which is diagonal, with elements $\mathrm{e}^{-E(x)/T}$. The time-reversal symmetry of the equilibrium ensemble of trajectories (5) corresponds to the operator equation

$$\mathbb{W}_{\mathrm{eq}}^{\dagger} = (\hat{\mathbb{T}}\hat{\pi})^{-1} \mathbb{W}_{\mathrm{eq}} (\hat{\mathbb{T}}\hat{\pi}) \tag{A.2}$$

This equation may be verified directly from the definitions of the various operators. (Note that $\hat{\mathbb{T}}^{-1} = \hat{\mathbb{T}}$, since it simply corresponds to a reversal of momenta.) We also introduce an operator $\hat{\mathbb{P}}$ corresponding to the spatial transformation $\mathbb{P}$, by taking $\hat{\mathbb{P}}|x\rangle = |\tilde{x}\rangle$. If the dynamics is invariant under $\mathbb{P}$, one has an operator equation

$$\hat{\mathbb{P}} \mathbb{W}_{\mathrm{eq}} \hat{\mathbb{P}} = \mathbb{W}_{\mathrm{eq}}. \tag{A.3}$$

(For the operator $\mathbb{W}_{\mathrm{eq}}$ in (A.1), this relationship is easily verified as long as $\partial E/\partial q_i$ is odd in $q_i$ for those co-ordinates $q_i$ which are inverted by $\mathbb{P}$.) We also note that $\int \mathrm{d}x \langle x|\mathbb{W}_{\mathrm{eq}}|P\rangle = 0$, independent of $|P\rangle$: this corresponds to conservation of probability.

To analyse the driven ensemble, we write $\mathbb{W}_{\rm neq} = \mathbb{W}_{\rm eq} - \sum_i f_i \frac{\partial}{\partial p_i}$ where the $f_i$ are the external forces (assumed independent of $p_i$). Since these forces are non-conservative, the relation (A.2) does not apply. However, the relation $\int \mathrm{d}x \langle x|\mathbb{W}_{\rm neq}|P\rangle = 0$ holds also for this non-equilbrium dynamics, since probability is (of course) still conserved.

To analyse the biased ensemble, we write $\mathbb{W}_{\rm bias}(\nu) = \mathbb{W}_{\rm eq} + \nu\hat{j}$ where the operator $\hat{j}$ is diagonal with elements $j(x)$. The theory associated with this operator is discussed in [11, 13, 38]. The operator $\mathbb{W}_{\rm bias}(\nu)$ does not have a probability-conservation property $\int \mathrm{d}x \langle x|\mathbb{W}_{\rm bias}(\nu)|P\rangle \neq 0$. However, the steady state probability distribution of $x$ in the biased ensemble is controlled by the largest eigenvalue of $\mathbb{W}_{\rm bias}$ and the associated left and right eigenvectors. Given the properties of the current discussed above (it is odd under both $\mathbb{T}$ and $\mathbb{P}$), then we have $\hat{\mathbb{P}}\hat{j}\hat{\mathbb{P}} = -\hat{j} = \hat{\mathbb{T}}\hat{j}\hat{\mathbb{T}}$. We also have $\hat{\pi}^{-1}\hat{j}\hat{\pi} = \hat{j}$, since these operators are all diagonal. Hence it follows from (A.2) that

$$\mathbb{W}_{\rm bias}(\nu)^\dagger = (\hat{\mathbb{P}}\hat{\mathbb{T}}\hat{\pi})^{-1}\mathbb{W}_{\rm bias}(\nu)(\hat{\mathbb{P}}\hat{\mathbb{T}}\hat{\pi}) \tag{A.4}$$

which is the promised operator equation corresponding to the symmetry (12).

To see the consequences of this equation, suppose that $\langle L|$ is the dominant left eigenvector of $\mathbb{W}(\nu)$ so that $|L\rangle$ is the dominant right eigenvector of $\mathbb{W}(\nu)^\dagger$. Then from (A.4) the dominant right eigenvector of $\mathbb{W}(\nu)$ is $|R\rangle = (\hat{\mathbb{P}}\hat{\mathbb{T}}\pi)|L\rangle$. The probability of configuration $x$ in the steady state is $P_{\rm bias}(x|\nu) \propto \langle L|x\rangle\langle x|R\rangle$ [38], so that $P_{\rm bias}(x|\nu) \propto L(x)L(\tilde{\bar{x}})\pi(x)$ where $\tilde{\bar{x}}$ is the phase space point obtained by applying $\mathbb{TP}$. Hence

$$P_{\rm bias}(\tilde{\bar{x}}|\nu) = P_{\rm bias}(x|\nu) \tag{A.5}$$

which is the symmetry relation for the steady state distribution of the biased process. Averages of one-time observables in the biased ensemble are fully determined by $P_{\rm bias}(x|\nu)$, so (A.5) specifies which quantities can have non-zero values in that ensemble, and which are constrained equal to zero by symmetry.

The strength of this operator approach is that the same algebraic structure can hold for a variety of different models. For example, there are many discrete Markov chain models where symmetries of the form (A.4) apply, including the simple symmetric exclusion process (SSEP) biased by the total current [3, 39, 40]. Thus, while we have concentrated throughout on systems with continuous co-ordinates $x = (\vec{q}, \vec{p})$, the operator formalism allows straightforward generalisations to overdamped Langevin dynamics (where $x = \vec{q}$) or to Markov chains such as the SSEP.

## Appendix B. Entropy production in biased ensembles, and the special case of a particle diffusing on a ring

As noted in Sec. 2.4, the definition of entropy production is slightly subtle in these biased ensembles. In that Section, we argue that the entropy production of a trajectory should be defined as $\Delta S(X) = \beta\Delta Q(X)$. If we consider a system at equilibrium (for which no work is done) then this implies [10] that

$$\Delta S(X) = \ln \frac{P_{\rm eq}[X|X_{-\tau}]}{P_{\rm eq}[\mathbb{T}X|(\mathbb{T}X)_{-\tau}]} \tag{B.1}$$

where $P_{\mathrm{eq}}[X|X_{-\tau}] = P_{\mathrm{eq}}[X]/p_0(X_{-\tau})$ is the probability of trajectory $X$, given that the system started in $X_{-\tau}$ at time $t = -\tau$. [To derive the relationship between trajectory probabilities and heat flow in this situation, it is sufficient to use the definition of $P_{\mathrm{eq}}[X|X_{-\tau}]$ in (B.1) and to note that since no work is done $\Delta Q(X)$ is equal to the energy change along the trajectory, which is equal to $T \ln \frac{p_0(X_\tau)}{p_0(X_{-\tau})}$.]

Now observe that given a trajectory $X$ of the system, the heat flow is a physically measurable quantity, and therefore depends on the trajectory $X$ but not on the bias $\nu$ (since the bias $\nu$ changes the probabilities of trajectories but has no effect on the trajectories themselves). Hence the entropy production in the biased ensemble is obtained by averaging $\Delta S(X)$ with respect to the distribution $P_{\mathrm{bias}}(X)$. Also note from (B.1) that $\Delta S(\mathbb{T}X) = -\Delta S(X)$, and that $\Delta S(\mathbb{P}X) = \Delta S(X)$. (The latter equality follows because all quantities in (B.1) are even under $\mathbb{P}$.) This means that $\Delta S$ is an observable of the second kind considered in Sec. 2.5, and its average must vanish.

However, an alternative definition of the entropy production would be to take

$$\Delta\tilde{S}(X) = \ln \frac{P_{\mathrm{bias}}[X|X_{-\tau}]}{P_{\mathrm{bias}}[\mathbb{T}X|(\mathbb{T}X)_{-\tau}]}. \tag{B.2}$$

This would be the entropy production that one would infer if, instead of measuring heat flow directly, one attempted to estimate the entropy production by direct inspection of the trajectory distribution $P_{\mathrm{bias}}$. In this case we would find

$$\Delta\tilde{S}(X) = 2\nu J(X) + \ln \frac{p_0\left(\overline{X}_\tau\right)}{p_0(X_{-\tau})} \tag{B.3}$$

where we used $P_{\mathrm{bias}}[X|X_{-\tau}] = P_{\mathrm{eq}}(X)\mathrm{e}^{\nu J(X)}/[p_0(X_{-\tau})Z(\nu)]$. This differs from the quantity $\Delta S(X)$ defined above by the term $2\nu J(X)$ whose average manifestly does not vanish in the biased ensemble, since $\langle J\rangle_{\mathrm{bias}} \neq 0$. Our conclusion in this paper – that biased ensembles are free from dissipation – is based on the result that $\langle \Delta Q\rangle_{\mathrm{bias}} = 0$, and is not affected by this ambiguity over the entropy production. However, to explore this issue in more detail, we now consider a simple model.

### *Appendix B.1. Diffusion on a ring*

Consider a single particle undergoing (overdamped) diffusive motion on a circle. Its co-ordinate is $x \in [0, 1)$ and it evolves according to

$$\partial_t x = -\nabla u(x) + \sqrt{2T}\eta + f \tag{B.4}$$

where $u(x)$ is a potential, $\eta$ is a white noise with covariance $\langle \eta(t)\eta(t')\rangle = \delta(t - t')$ and $f$ is a constant applied force that drives the particle around the circle. We assume periodic boundary conditions and $u(x)$ is a periodic function $u(x) = u(1 + x)$.

The case $f = 0$ represents the equilibrium system while the special case $u = 0$ corresponds to free diffusive motion in the absence of any potential. If $u = 0$ and $\nu = 0$, it is easy to prove that the only effect of the force $f$ is to bias the diffusive motion, so if the system is at position $x'$ at time $t = 0$ then the distribution of its position a time $t$ later is

$$G_f(x, t; x', 0) = \frac{1}{\sqrt{4\pi T t}} \mathrm{e}^{-(x-x'-ft)^2/4Tt} \tag{B.5}$$

On the other hand, if $u = 0$ and $f = 0$ but we consider the biased ensemble (with $\nu \neq 0$ and $J = \int (\partial_t x) \mathrm{d}t$ and $\tau \to \infty$) then the corresponding probability is [28]

$$G_\nu(x, t; x', 0) = \frac{1}{\sqrt{4\pi T t}} \mathrm{e}^{-(x-x'-2\nu t)^2/4Tt} \tag{B.6}$$

Since all information about the system is encoded in this two-time correlation function, it follows for $f = 2\nu$ that the biased and driven ensembles lead to the same behaviour in this case.

In fact, this system (with $u = 0$) is a special case because *this driven system has a PT symmetry*, which is not typical of driven systems in general. To see this, note that the statistical properties of a random walk in which a particle hops preferentially to the right are exactly the same as those obtained by time-reversing a random walk in which the same particle hops preferentially to the left (by the same amount). This can be verified directly, or by noting that the generator of the biased random walk satisfies (A.4). However, it is also easily verified that if $\nabla u \neq 0$ and $f \neq 0$ then the generator of the process (B.4) does not satisfy (A.4) so the driven system has no PT symmetry. This latter situation, in which the driven and biased systems are qualitatively different, is the more general one.

### *Appendix B.2. Physical signatures of dissipation*

To understand the physical significance of the special case ($u = 0$) where the driven particle diffusing on a ring has a PT symmetry, it is useful to consider a physical setting in which (B.4) might apply. Imagine a colloidal particle immersed in a stationary solvent. This particle is localised (for example by optical tweezers) to lie in a circular region of space, and then the same tweezers could be employed to drive the particle around the circle [42]. In this case the tweezers would do work on the system and this work would be dissipated as heat in the solvent.

Now consider the same particle, still localised to lie on a circle. To construct the biased ensemble, the optical tweezers are *not* used to drive the particle around the circle. Instead, one waits for spontaneous events to cause the particle to exhibit a mean current in the clockwise direction. Using (B.4) to model the motion of the colloidal particle, this problem can be solved exactly [28]. One finds that motion of the colloidal particle in this case is the same as when it is driven by the tweezers. That is, if we inspect only the statistics of the motion of the colloidal particle, the rare spontaneous events (without applied force) are indistinguishable from those where a driving force is applied. As we explained above, this is a special case, since the driven system has a PT symmetry only if $u = 0$: in the general case $u \neq 0$ then the spontaneous fluctuations *can* be distinguished from the driven systems since only the spontaneous fluctuations have PT symmetry.

Moreoever, if we consider the physical system of interest instead of the simplified model (B.4), we find that the driven process and the biased process can be distinguished, as long as we observe the solvent properties as well as the colloid. When the force is applied to the colloid, we expect local heating of the solvent due to the work done by the tweezers, but when we observe rare events, there is no such heating. The case $u = 0$ is peculiar because it is not possible to detect this heat flow by inspection of the colloidal particle alone. In this sense, the coarse-grained model (B.4) that we are using to describe the system is insufficient to capture the dissipation (and associated entropy production) that is taking place in the driven system. Again, this is a feature of the special case $u = 0$ in this one-particle system.

The dependence of entropy production on the environment in which an object is moving has been discussed previously [43, 44]. For the purposes of this work, we believe that the ambiguity over the definition of the entropy production in the biased ensemble ($\Delta S$ or $\Delta\tilde{S}$) is based on a similar effect, which is related to the question of whether an external force is doing work on the colloid, or whether some rare realisation of random solvent collisions causes the colloid to move. While this may appear to be a semantic question, we argue here that these two cases can be distinguished (typically), because the first case (usually) results in breaking of PT-symmetry, but the second does not.

## References

[1] H. Touchette, Phys. Rep. **478**, 1, (2009).
[2] D. Ruelle, *Thermodynamic Formalism* (Addison-Wesley, Reading, 1978); J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[3] T. Bodineau and B. Derrida, Phys. Rev. Lett. **92**, 180601 (2004).
[4] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio and C. Landim, J. Stat. Phys. **135**, 857 (2009).
[5] B. Derrida, J. Stat. Mech. 2007, P07023 (2007)
[6] G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. **80**, 931 (1995).
[7] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[8] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[9] C. Maes, J. Stat. Phys. **95**, 367 (1999).
[10] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[11] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[12] V. Lecomte, C. Appert-Roland and F. van Wijland, Phys. Rev. Lett. **95**, 010601 (2005).
[13] V. Lecomte, C. Appert-Roland and F. van Wijland, J. Stat. Phys. **127**, 51 (2007).
[14] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk and F. van Wijland, Phys. Rev. Lett. **98**, 195702 (2007).
[15] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk and F. van Wijland, J. Phys. A **42**, 075007 (2009).
[16] L. O. Hedges, R. L. Jack, J. P. Garrahan and D. Chandler, Science **323**, 1309 (2009).
[17] R. M. L. Evans, Phys. Rev. Lett. **92**, 150601 (2004).
[18] R. M. L. Evans, J. Phys. A: Math. Gen. **38**, 293 (2005).
[19] A. Baule and R. M. L. Evans, Phys. Rev. Lett. **101**, 240601 (2008).

[20] M. Knežević and R. M. L. Evans, Phys. Rev. E **89**, 012132 (2014).
[21] R. M. L. Evans, R. A. Simha, A. Baule, and P. D. Olmsted, Phys. Rev. E **81**, 051109 (2010).
[22] C. Maes and M. H. van Wieren, Phys. Rev. Lett. **96**, 240601 (2006).
[23] M. Baesi, C. Maes and B. Wynants, Phys. Rev. Lett. **103**, 010602 (2009).
[24] E. T. Jaynes, Annu. Rev. Phys. Chem. **31**, 579, (1980).
[25] S. Pressé, K Ghosh, J. Lee and K. A. Dill, Rev. Mod. Phys. **85**, 1115 (2013).
[26] M. Otten and G. Stock, J. Chem. Phys. **133**, 034119 (2010).
[27] R. Chétrite and H. Touchette, Ann. Henri Poincaré **16**, 2005 (2015).
[28] R. Chétrite and H. Touchette, Phys. Rev. Lett. **111**, 120601 (2013).
[29] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, (Academic Press, London, 2006), chapter 7.
[30] A. Simha, R. M. L. Evans, and A. Baule, Phys. Rev. E **77**, 031117 (2008).
[31] R. M. L. Evans, C. A. Hall, R. A. Simha, and T. S. Welsh, Phys. Rev. Lett. **114**, 138301 (2015).
[32] G. Bussi and M. Parrinello, Phys. Rev. E **75**, 056707 (2007).
[33] D. A. Sivak, J. D. Chodera and G. E. Crooks, Phys. Rev. X **3**, 011007 (2013).
[34] P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, Ann. Rev. Phys. Chem. **53**, 291 (2002).
[35] T. Speck and D. Chandler, J. Chem. Phys. **136**, 184509 (2012)
[36] C. J. Fullerton and R. L. Jack, J. Chem. Phys. **138**, 224506 (2013).
[37] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).
[38] R. L. Jack and P. Sollich, Prog. Theor. Phys. Supp. **184**, 304 (2010).
[39] C. Appert-Rolland, B. Derrida, V. Lecomte and F. van Wijland, Phys. Rev. E **78**, 021122 (2008).
[40] R. L. Jack, I. R. Thompson and P. Sollich, Phys. Rev. Lett. **114**, 060601 (2015).
[41] R. L. Jack and P. Sollich, Eur. Phys. J.: Special Topics **224**, 2351 (2015).
[42] V. Blickle, T. Speck, C. Lutz, U. Seifert and C. Bechinger, Phys. Rev. Lett. **98**, 210601 (2007)
[43] R. A. Blythe, Phys. Rev. Lett. **100**, 010601 (2008)
[44] T. Speck, J. Mehl and U. Seifert, Phys. Rev. Lett. **100**, 178302 (2008)