nature reviews physics

https://doi.org/10.1038/s42254-024-00720-5

Review article

Check for updates

Entropy, irreversibility and inference at the foundations of statistical physics

Jonathan Asher Pachter [1,2], Ying-Jen Yang [1] & Ken A. Dill [1,2]

Abstract

Statistical physics relates the properties of macroscale systems to the distributions of their microscale agents. Its central tool has been the maximization of entropy, an equilibrium variational principle. Recent work has sought extensions to non-equilibria: across processes of change both fast and slow, in the Jarzynski equality and fluctuation relations and other tools of stochastic thermodynamics, using large deviation theory or others. When recognized as an inference principle, entropy maximization can be generalized for non-equilibria and applied to path entropies rather than state entropies, becoming the principle of maximum caliber, which we emphasize in this Review. Our primary goal is to enhance crosstalk among researchers working in disparate silos, comparing and contrasting different approaches while pointing to common roots.

Sections

[1]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY, USA. [2]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA. e-mail: dill@laufercenter.org

Key points

- The original concept of equilibrium thermodynamic entropy has branched into two related but distinct concepts, both termed entropy: one a tool for inference and the other a measure of time irreversibility.
- The field of stochastic thermodynamics and the methods therein have developed the irreversibility version of entropy, extracting many important results, although some of those results are limited in interpretation to near equilibrium.
- Although many problems in statistical physics can be formulated in terms of multiple approaches, the Bayesian inferential approach provides the most general and solid footing.
- In analogy with the maximum entropy inference approach to equilibrium thermodynamic states, the maximum caliber principle performs inference with path entropies, serving as a powerful generative procedure for making models in statistical physics and beyond.

Introduction

Statistical physics emerged in the mid-1800s as a way to explain and predict macroscopic material phenomena from the physics of their constituent atoms, molecules or other particles. The basic methods are well known and given in textbooks[1–5]. They often reason from the Gibbs ensemble method and are mostly applied to molecules and materials, at or near equilibrium, exchanging with idealized baths. But in recent years, statistical physics has been applied to biology[6], ecology[7], economics[8], active matter[9], artificial intelligence and neural networks[10] with foundational developments in stochastic thermodynamics[11,12], the Jarzynski equality[13,14], macroscopic fluctuation theory[15], large deviation theory (LDT)[16] and maximum caliber (MaxCal)[17,18]. This Review grew out of a need for a comprehensive roadmap that communicates the different premises, purposes and limitations of these core ideas, across different processes of change, and yet linked through the common language of history and principles. We assume some background in undergraduate-level probability theory, thermodynamics and statistical physics[1–5].

We first note two disparate perspectives used in statistical physics. From its earliest days, statistical physics was framed in terms of Hamilton's equations of the positions and momenta of collisional particles, their conservation of energy through the Liouville equation and puzzles of irreversibility and disorder[19–21]. The second perspective — which is our main focus here — is the language of probabilities, the basis for most present-day problem-solving. The probabilistic basis extends the reach of statistical physics to problems and processes well beyond collisional particles experiencing temperature and pressure. Newtonian mechanics is an unnecessarily limiting starting point for developing the most general principles of model-making, especially for complex, possibly living, systems whose constituents are themselves high dimensional[22]. Furthermore, the probability framework gives tools that are practical and simple to apply, with no sacrifice in rigour.

In this Review, we briefly recall the origins of entropy in thermodynamics, tracing the different justifications that have historically been used for maximizing entropy. This leads to the modern inferential view of entropy maximization as the unique procedure for generating statistical models that both obeys internal logical consistency and introduces no unwarranted bias into the description of a system. We then contrast this inferential viewpoint with the use of entropy to measure time irreversibility in a non-equilibrium process, as exemplified by the field of stochastic thermodynamics. Although this irreversibility viewpoint provides powerful results for characterizing non-equilibrium processes, we again emphasize that it is most broadly impactful to use entropy as a tool for inference. Along these lines, we conclude by outlining how maximizing the caliber — an entropy over possible paths — can have the same power to describe non-equilibrium situations as the traditional maximum entropy approach has had in the world of equilibrium statistical physics.

A tale of two entropies at the heart of equilibrium statistical physics

At the heart of equilibrium statistical physics is a remarkable equality between two seemingly unrelated quantities — a macroscopic physical property of matter and a mathematical property of probability distributions — which are both now called entropy:

$$\Delta S_{\text{Clausius}} = \Delta S^{*}_{\text{BG}}. \tag{1}$$

On the left side of equation (1) is the change in Clausius entropy, which encapsulates certain transformations between thermodynamic equilibria. On the right side is the change in maximized Boltzmann–Gibbs (BG) entropy, a mathematical property of a probability distribution over microscopic degrees of freedom of a system. In the following sections, we define these two entropies and review the basis for the principle embodied in equation (1).

The Clausius entropy describes macroscopic processes

At the foundation of thermodynamics is an inequality owing to Rudolf Clausius, accompanied by a definition of reversible, quasi-equilibrium processes as a special limiting case of dynamics. Clausius' theorem states that the integral of the infinitesimal heat δQ a system dissipates to its environment, divided by temperature T, is always non-negative along a cyclic transformation and is only zero for a reversible cyclic process (this is a quantitative counterpart to qualitative statements of the second law of thermodynamics, such as the Kelvin formulation describing the impossibility of spontaneous heat flow from a cooler object to a hotter object[1]):

$$\oint_{\text{general}} \frac{\delta Q}{T} \geq \oint_{\text{reversible}} \frac{\delta Q}{T} = 0. \tag{2}$$

This result can be used to define the change in Clausius (or thermodynamic) entropy as this heat integral along any reversible trajectory, which must be less than or equal to the same integral for a general trajectory connecting the same two thermal equilibrium states, owing to the theorem mentioned earlier:

$$-\Delta S_{\text{Clausius}} = \int_{\text{reversible}} \frac{\delta Q}{T} \leq \int_{\text{general}} \frac{\delta Q}{T}. \tag{3}$$

Although $\Delta S_{\text{Clausius}}$ can rationalize why heat flows from hotter to colder states, it contains no information about time dependence, such as how long two bodies at different temperatures would take to equilibrate with each other. It also says nothing about the underlying microscopic constituents of matter.

Statistical physics arose from the need for probability distributions

The need for probability distributions in connecting macroscale to microscale behaviour was already made evident by the development of the kinetic theory of gases[23]. Although the mean velocity v of a particle of mass μ in a gas at equilibrium is zero (and therefore uninformative), the variance of velocity was found to be related to the temperature through $\langle v^2 \rangle = 3k_B T/\mu$ for an ideal gas, in which k_B is the Boltzmann's constant. Where there is a variance, there is uncertainty and there is a need for distributions over microscopic agents to explain some macroscopic behaviours.

Throughout its history, statistical physics has been enmeshed in a broader controversy about the nature of probabilities themselves. Probabilities are mathematically defined as fractional quantities, associated with events that can be combined into Boolean collections of independent and mutually exclusive events, obeying the addition and multiplication logical rules of probability[24]. The controversy is in whether probabilities can only be descriptions of frequencies of replicable events, as when dice are rolled multiple times (a frequentist view), or can more broadly also describe inferences about non-repeatable events (a Bayesian view). Although the methodology of statistical physics has been stable since its earliest days, how that methodology is justified on principle has changed — initially primarily as frequentist, but with the Bayesian inferential view gaining more traction over the decades.

Probabilities interpreted as frequencies in Gibbs' ensembles

The Gibbs framework for statistical physics, which followed from preliminary work by others[25], casts the theory in terms of 'ensembles' and employs frequencies, the prevailing language for probabilities at the time it was developed. The concept of an ensemble is simple. Consider a system with M possible microstates, such as M possible ways of arranging its microscopic constituents. Now, imagine there are N identical copies of the system, each subject to the exact same macroscopic (or thermodynamic) external conditions. One can count the number of times n_m that microstate m occurs within this ensemble of copies, giving a set of occurrence numbers for all the different m. The number of different ways one could observe the same set of occurrences $\{n_m | m = 1, 2, \ldots, M\}$ with $\sum_m n_m = N$ is a multinomial:

$$W = N! \Big/ \left(\prod_{m=1}^{M} n_m! \right). \tag{4}$$

Defining the frequency of each outcome as $f_m = n_m/N$, the (intensive) Boltzmann definition of entropy can be calculated in terms of multiplicity $S = k_B \ln W/N$. In the limit of an infinitely large ensemble $N \to \infty$, applying the Stirling approximation for factorials yields the BG entropy:

$$S_{BG}[f_m] = \lim_{N\to\infty} \frac{k_B \ln W}{N} = -k_B \sum_{m=1}^{M} f_m \ln f_m . \tag{5}$$

A mechanical model of states includes conserved quantities, such as a delineation of the energy ε_m for each microstate m. For example, in the Maxwell–Boltzmann treatment of an ideal gas, the microstates are the velocities of a particle of mass μ, and the energy of a microstate is $\varepsilon_v = \mu v^2/2$. In the standard description of canonical ensembles, the Boltzmann distribution for a system in a heat bath is derived by minimizing the Helmholtz free energy F:

$$F = \sum_{m=1}^{M} f_m \varepsilon_m - TS_{BG}, \tag{6}$$

in which $\sum_{m=1}^{M} f_m \varepsilon_m$ is the average energy. Minimizing F is the same as maximizing

$$-\beta F = -\sum_{m=1}^{M} f_m \ln f_m - \beta \sum_{m=1}^{M} f_m \varepsilon_m, \tag{7}$$

in which $\beta = 1/k_B T$, which is tantamount to maximizing the BG entropy S_{BG} with the average energy treated as a constraint with a Lagrange multiplier β. The frequency that minimizes F is the Boltzmann distribution

$$f_m^* = \frac{e^{-\beta\varepsilon_m}}{Z}, \tag{8}$$

in which

$$Z = \sum_{m=1}^{M} e^{-\beta\varepsilon_m} \tag{9}$$

is the partition function, a central quantity in equilibrium statistical physics, related to the free energy F by $\ln Z = -\beta F$.

LDT and some problems of data inference

We pause here to introduce LDT. Although the most studied aspects of probability distributions are often the low-order moments, that is, the means and variances, nevertheless, the rare events and the tails of distributions are often of interest too. LDT has applications in statistical physics[16], but it is particularly relevant in what could be called 'data-rich' situations, wherein events are replicable, leading to the applications of Gibbs ensemble ideas to any 'idealized infinite data set'[22,26].

LDT formalizes the connection, described earlier for Gibbs ensembles, that the state of maximum multiplicity is that which is most likely observed. Assuming a uniform prior distribution (according to the principle of equal a priori probabilities[5]), the probability of observing a particular set of frequencies $\{f_m\}$ in the Gibbs ensemble is proportional to the multiplicity W and thus has the following asymptotic form[27]:

$$\mathrm{Prob}(f) = \exp\left[-N \sum_{m=1}^{M} f_m \ln f_m + o(N) \right]. \tag{10}$$

$o(N)$ are terms that are much smaller than N in the limit of large N. When $N \to \infty$, this probability distribution is exponentially dominated by the frequency that is closest to the uniform prior, as solved by maximizing S_{BG}, which appears in the expression inside the exponential in equation (10). Making a quadratic approximation of S_{BG} for small, linear deviations leads to a Gaussian distribution[28], as dictated by the central limit theorem. But equation (10) also contains information about f for larger deviations. When data are introduced that constrain the possible frequencies, then the most probable frequency f^* under the constraints is the one that uniquely maximizes S_{BG}. In such a 'data-rich' situation, entropy maximization in Gibbs' framework is equivalent to Bayesian model updating[22,26]. This LDT formulation is useful in generalizing Gibbs' framework to infer Markov processes from correlated data[22,29,30].

The equality of the two entropies

We return to equation (1), the equivalence between two entropies. Substituting the posterior f_m^* of equation (8) into equation (7) for the BG entropy yields the maximized entropy $S_{\mathrm{BG}}^*/k_{\mathrm{B}} = \beta\langle\varepsilon\rangle + \ln Z$. Consider a process in which the system is transformed by a perturbation to the energy levels of $\delta\varepsilon_m$. The change in $\ln Z$ is

$$\begin{aligned}\delta \ln Z &= \sum_{m=1}^{M} \frac{1}{Z}\frac{\partial Z}{\partial \varepsilon_m}\delta\varepsilon_m \\ &= -\beta \sum_{m=1}^{M} \frac{1}{Z} e^{-\beta\varepsilon_m}\delta\varepsilon_m \\ &= -\beta \sum_{m=1}^{M} f_m^*\,\delta\varepsilon_m \\ &= -\beta\delta W,\end{aligned} \tag{11}$$

in which δW is the work performed on the system in a reversible process. Thus the change in maximized BG entropy is

$$\delta S_{\mathrm{BG}}^*/k_{\mathrm{B}} = \beta\delta\langle\varepsilon\rangle + \delta \ln Z = \beta\delta(\langle\varepsilon\rangle - W) = -\beta\delta Q = \delta S_{\mathrm{Clausius}}/k_{\mathrm{B}}, \tag{12}$$

in which δQ is the heat dissipated in this reversible process.

The ensemble approach is useful for predicting macroscopic properties of microscopic models. However, its conceptualization of probabilities as frequencies limits its applicability beyond replicable events. The ensemble approach envisions a set of infinite copies of a system, but in reality there is often only one system. The replicates could arise from different snapshots of a system spaced out in time. This conception, however, requires two companion assumptions that restrict its generality: first, the ergodic hypothesis, which equates time averages with ensemble averages, and second, the independence of replicates due to chaotic interactions or random collisions. These limitations can be circumvented by framing statistical physics instead as a matter of drawing inferences in probabilistic modelling.

Entropy maximization and consistent inference in model-making

In the inference-based perspective, the maximum entropy principle (MaxEnt) is the unique procedure for generating probability distributions that are logically consistent (obeying the laws of probability) and consistent with known, externally supplied information. In this view, probabilities are not limited to being frequencies of repeatable events and instead can characterize uncertainties of all kinds. In this Bayesian school of thought, one can discuss the probability of a nuclear disaster, life on Mars, or the chance of snow at some particular time and date in the future, even though none of these are repeatable events. Reframing the foundations of statistical physics in terms of Bayesian inference avoids the need for replicates of physical systems; efforts at such re-framing date to work from Edwin Thompson Jaynes in the 1950s[31].

Jaynes saw statistical physics as a procedure for inference and model-making rather than as a law of nature. He elevated the MaxEnt calculation to inferring a probability model p_m instead of a set of frequencies f_m. In the standard textbook example described earlier in equations (5)–(7), the framework of Jaynes infers the posterior model p_m^* by maximizing $-\beta F$ of a probability distribution p_m:

$$-\beta F = -\sum_{m=1}^{M} p_m \ln p_m - \beta \sum_{m=1}^{M} p_m \varepsilon_m. \tag{13}$$

Although equation (13) is mathematically equivalent to equation (7), its basis in inference gives it a deeper logical foundation and broader applicability, as noted earlier. Jaynes framed it as a pipeline of model inference: asserting only the information that one has (for example, average energy $\langle\varepsilon\rangle$ in the paradigmatic calculation mentioned earlier) and asserting maximal ignorance (quantified by entropy) otherwise. Arguments of Jaynes removed the dependence on replicability of the event, but some physicists objected that statistical physics should not be tied to the state of knowledge of an observer[32], as implied by the use of the term 'ignorance'.

This conceptual problem is resolved by fully axiomatizing the inference process[33]. Given a prior distribution q_m representing some initial information about the system, and given new information constraining the possible forms for the posterior distribution p_m, there is only one posterior distribution p_m^* obeying consistency with the logical structure of probabilities which introduces no unwarranted bias; the unique posterior p_m^* is what maximizes the BG relative entropy R_{BG} of the posterior with respect to the prior q_m, constrained by the relevant information:

$$R_{\mathrm{BG}} = -\sum_m p_m \ln\left(\frac{p_m}{q_m}\right). \tag{14}$$

Although R_{BG} can be computed for all possible posteriors with respect to the given prior, it is only the maximum value R_{BG}^* that selects the proper posterior p_m^*. Up to a constant, R_{BG} reduces to the original BG entropy S_{BG} described earlier in equations (5)–(7) when the prior is uniform. The common assumption of a uniform prior, known as the principle of equal a priori probabilities, is sometimes justified by symmetry arguments. For example, if the microstates are positions of a particle, and the system has isotropic symmetry, then the starting point of knowledge is that no position should be different than any other position, and thus all positions are equally likely.

In this inference-based perspective, statistical physics entails (Fig. 1): a user-defined space of possible microstates, a user-defined prior distribution over the microstates, appropriate user-defined constraints, such as known averages from external data, and the MaxEnt procedure, predicting the posterior distribution that faithfully represents the premises and prior without unwarranted bias. In this view, MaxEnt is not a law of physics; rather, MaxEnt is a method of drawing inferences. A user asserts a particular model (that is, an indexed state space, usually its dependence on some physical variable or variables, a prior and constraints) and computes a posterior distribution. If a predicted distribution is found to be inconsistent with experiments, the burden is on the user to find a better model, because the procedure itself is built upon rigour.

'Non-extensive' statistical physics

MaxEnt typically yields exponential distributions. However, various systems that are ubiquitous in nature display non-exponential distributions, such as those having power-law tails[34]. What approach would predict non-exponential distributions? A field called non-extensive statistical physics[35] derives non-exponential distributions using non-BG entropies, such as Tsallis or Renyi entropies. However, the arguments outlined in the previous section regarding consistent inference show that non-BG entropies introduce unwarranted bias[36–38]. The relevant axiom for consistent inference in the discussion here is what has been called strong-system independence[39], which has been summarized as the combination of subset independence and minimal updating[40]:

when two systems are a priori assumed to be independent and new information does not reveal whether the systems are independent or not, the independence in the prior should be maintained. Proper inference without unwarranted bias can only be done with the BG entropy and its maximization.

BG inference is not limited to predicting exponentials. Non-exponential distributions are predicted through BG inference when correlations are properly accounted for in the imposed constraints. For example, in social and financial situations, an 'economy of scale' model gives non-exponential distributions with power-law tails[41]. Exponentials arise when the function whose expectation is being constrained — which in classic statistical physics systems is the energy ε_m — is any function $f(m)$ of only the state index m. However, if the energy or energy-like function $\varepsilon_m = g(m, \varepsilon_n)$ is a function both of state index m and of the energy values of other states[41], power-law tails emerge naturally from the maximization of BG entropy. Other methods that yield non-exponentials from BG entropy include superstatistics (which itself is simply a nested MaxEnt procedure)[42–45] and non-ideal reservoirs[46–48], among others[49–52].

Defining irreversibility beyond equilibria

We now turn to non-equilibrium statistical physics. We first ask: can various forms of irreversibility in nature be quantified by some entropy-like variational principle? Here, we summarize the ideas underlying Boltzmann's H-theorem, the Jarzynski equality, stochastic thermodynamics and macroscopic fluctuation theory. These works are advances beyond the Clausius restrictions to slow, reversible processes between two equilibrium states. But, as described subsequently, they are not fully free of restrictions, in some cases requiring equilibrium baths and an assumption of local detailed balance (LDB).

Boltzmann's H-theorem and irreversible relaxations

Links between microscopic particles and macroscopic thermodynamics can be found via the following entropy-like functional of the distribution density $f(x, v, t)$ for particles in a gas to have position x and velocity v at time t:

$$H = -\int\int \mathrm{d}v\mathrm{d}x\, f(x, v, t) \ln f(x, v, t), \tag{15}$$

which reduces to the Boltzmann entropy $\ln W$ when all microstates are equally probable.

This quantity is used to state Boltzmann's H-theorem: namely, that H increases in time until the gas equilibrates, at which point H attains its maximum value equal to the equilibrium thermodynamic entropy[1] (Fig. 2). H is an early example of a Lyapunov function — a quantity that evolves monotonically in time for a particular dynamics, thus characterizing the irreversible relaxation to steady state.

For Markov dynamics that have a steady state, a well-known Lyapunov function is the Kullback–Leibler (KL) divergence[53]:

$$D(t) = D(p(t)||\pi) = \sum_i p_i(t) \ln \frac{p_i(t)}{\pi_i}, \tag{16}$$

in which $p_i(t)$ is the probability distribution at time t and π_i is the steady-state probability distribution, with the double bar denoting that this is a measure of divergence of the distribution $p(t)$ relative to the distribution π (in previous sections, we have used p to denote a steady-state distribution, such as a Boltzmann distribution; here we use $p(t)$ to denote time-dependent distributions and we use π specifically for the steady state).

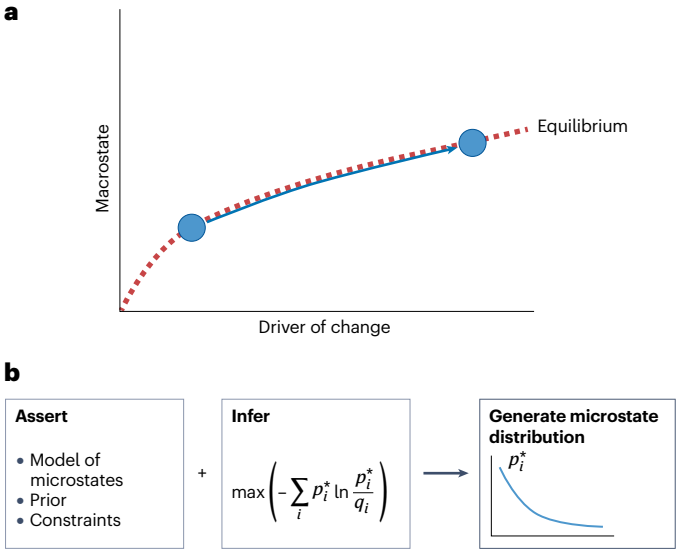

Fig. 1 | The generative inference pipeline of equilibrium statistical physics, which generates the microscopic probability distribution from the input assertions. a, 'Driver of change' refers to whatever external variable, such as temperature or pressure, is altered to bring a system from one equilibrium macrostate to another. **b**, The modeller asserts microstates and accompanying constraints in terms of those microstates, along with a prior distribution q_i; the MaxEnt procedure then infers the unique probability distribution p_i^* that satisfies the constraints (such as data) and the laws of probability.

If the dynamics has a unique steady state, then it is always the case that $\mathrm{d}D(t)/\mathrm{d}t \leq 0$; the monotonic decrease of $D(t)$ characterizes the stability of the steady state[54] and the magnitude of the change in $D(t)$ as a system relaxes to equilibrium is equal to the maximum possible work one could extract from the relaxation of the system[55]. The field of information geometry[56] provides additional insights along these lines, for example, giving an upper limit on the relaxation speed of an observable related to its fluctuations and the local curvature of $D(t)$[57–59].

For dissipative deterministic dynamics, a Lyapunov function known as the quasi-potential can be derived by considering the deterministic dynamics as the zero-noise limit of a stochastic dynamics[60]. This Lyapunov function is useful in analysing limit cycles[61], Kramers' theory of barrier-crossing in chemical reaction kinetics[60,62] and non-equilibrium dynamics in living systems in general[63–65]. More recently, the quasi-potential has been given a stochastic thermodynamics interpretation[66,67], unifying it conceptually with the stochastic Lyapunov function described earlier.

Stochastic thermodynamics computes properties of irreversibility

The field of stochastic thermodynamics[14,68,69] has emerged in the past few decades, generalizing Clausius' theorem to more broadly quantify irreversibility. It inherits key concepts from chemical thermodynamics[70] and from chemical-reaction network physics[54].

Note that equation (3) admits a separation into two components for any process starting and ending in equilibrium:

$$\int_{\text{general}} \frac{\delta Q}{T} = -\Delta S_{\text{Clausius}} + I, \tag{17}$$

in which $-\Delta S_{\text{Clausius}}$ is this heat integral calculated along a reversible trajectory, and I is the remainder, corresponding to the irreversibility of the process.

Stochastic thermodynamics seeks a relatively similar parsing into two components — one corresponding to an idealized reversible process and another corresponding to the irreversibility — but for Markov processes described by rate coefficients k_{ij} for transition from state i to j that parameterize the master equation. At the heart of stochastic thermodynamics is the quantity δA_{ij} called affinity, associated with each transition, defined as

$$\delta A_{ij} = \ln\left(\frac{k_{ij}}{k_{ji}}\right). \tag{18}$$

The necessary and sufficient condition for a Markov process to be detailed balanced is for the affinity to add up to zero around any cycles. This is the Markov process analogue of Clausius' theorem and in mathematics is also known as the Kolmogorov criterion[71,72]. In analogy with the Clausius picture, the affinity can be decomposed into an equilibrium-like landscape-descending term and an additional term that quantifies the breakdown of detailed balance:

$$\delta A_{ij} = \ln\left(\frac{k_{ij}}{k_{ji}}\right) = -\Delta(-\ln\pi) + \ln\left(\frac{\pi_i k_{ij}}{\pi_j k_{ji}}\right), \tag{19}$$

in which π_i is the steady-state distribution. Taking the dynamical average of equation (19) — weighted by the trajectories, not states — gives the Markov-process parsing equivalent of equation (17):

$$\langle \delta A\rangle = \sum_{i<j} J_{ij} \ln\left(\frac{k_{ij}}{k_{ji}}\right) = -\Delta\langle -\ln\pi\rangle + \underbrace{\sum_{i<j} J_{ij} \ln\left(\frac{\pi_i k_{ij}}{\pi_j k_{ji}}\right)}_{\geq 0} \tag{20}$$

in which $J_{ij} = \pi_i k_{ij} - \pi_j k_{ji}$ is the steady-state probability flux.

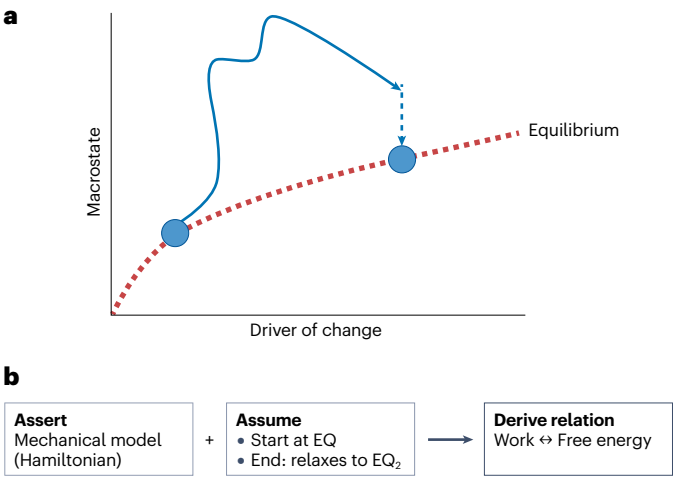

Fig. 2 | Hamiltonian-constrained non-equilibrium modelling. a, Endstates are equilibrated with ideal baths, but the routes (blue line) can be arbitrarily far from equilibrium because they obey Hamiltonian dynamics. Examples are the Jarzynski equality and Boltzmann H-theorem, the latter of which describes a relaxation to equilibrium (the downward dashed arrow). **b**, These approaches are not generative of models; rather, they assume Hamiltonian dynamics and a Boltzmann distribution at at least one endstate, to generalize work and free energy relations beyond the Clausius inequality.

The affinity decomposition of equation (19) exemplifies the key idea of stochastic thermodynamics, inherited from Clausius: dynamical properties can be split into a detailed-balance component and a non-balanced component; the latter measures how far the system deviates from equilibrium. This type of decomposition occurs throughout the literature, with the arguably best-known example[73–75] being the parsing of Markov irreversibility $I_{\text{tot}}(t)$ into deviation from its steady state (non-stationarity), quantified by the KL divergence of equation (16) $-\mathrm{d}D(t)/\mathrm{d}t \geq 0$; and an additional irreversibility owing to detailed-balance breakdown (DBB):

$$I_{\text{DBB}}(t) = \sum_{i<j} J_{ij} \ln\left(\frac{\pi_i k_{ij}}{\pi_j k_{ij}}\right) \geq 0. \tag{21}$$

The total irreversibility rate in stochastic thermodynamics, the sum of these two terms, is the KL divergence of the probability of a two-state trajectory with respect to its time reversal:

$$I_{\text{tot}}(t) = \sum_{i,j} p_i(t) k_{ij} \ln\left(\frac{p_i(t)k_{ij}}{p_j(t)k_{ji}}\right) = -\frac{\mathrm{d}D}{\mathrm{d}t}(t) + I_{\text{DBB}}(t). \tag{22}$$

In short, for detailed-balanced systems, irreversibility comes solely from the relaxation to the steady state. But for non-balanced systems, there is non-zero probability flux even at steady state, producing a non-zero I_{DBB}, which captures the irreversibility owing to an underlying non-equilibrium steady state.

Consider these three irreversibilities in stochastic dynamics: I_{tot}, and its components I_{DBB} and $-\mathrm{d}D/\mathrm{d}t$. Each can be rewritten as a KL divergence between the path probability of a forward process and a reversed process[74–77]. This path-entropic formulation has led to a zoo of fluctuation relations for stochastic thermodynamics of Markov processes[14,76,77] and can help identify the domains of validity for different fluctuation relations[77]. Additional stochastic thermodynamics results relate irreversibility and flux, such as the thermodynamic uncertainty relation that bounds the Fano factor of flux by time irreversibility[78,79]. These relations describe how tuning up time irreversibility enables — or quantitatively improves — various biological functions, such as kinetic proofreading[80,81], sensory adaptation[82] and directed motion of a molecular motor[83,84].

The same decomposition technique proves useful in analysing other quantities. In a generalized Einstein relation, the detailed-balanced component of the covariance between the state of the system and its rate of change is related to fluctuations[85]. Another example is the velocity and flux decomposition for diffusion processes[67,73,86,87] and its zero-noise limit[60,67]. Velocity and flux in these systems are decomposed into gradient-descending parts (the detailed-balance component) and a complementary cyclic part embodying the breaking of detailed balance; this is the method for finding a Lyapunov landscape function that we noted earlier[60,66,67]. In short, non-balanced dynamical systems require more than mere scalar landscape terms to describe them. In fact, the so-called dynamical terms or curl-flux terms[67,73] can have important implications in biology[64]. These terms can explain how biological cell development is asymmetrical: cells can generally differentiate in growing organisms but not de-differentiate[64].

Relating dynamics to thermal physics via the LDB assumption

It is often of interest to relate the time irreversibility of a process, through its transition rates k_{ij}, to energetics, which typically entails assuming LDB. LDB asserts that a given process has an idealized

timescale separation such that the fast modes are fast enough to achieve a local equilibrium before the slow modes have any noticeable changes. Then, one can coarse grain over the fast modes to focus on states corresponding to the slow modes (typically called mesostates in this situation) and assign a free energy to each mesostate because the fast modes have achieved the local equilibrium within each mesostate. Denoting by ΔF_{ij} the difference in locally defined free energy landscape between two mesostates, this procedure leads to a relation for the transition rates k_{ij} between mesostates

$$\frac{k_{ij}}{k_{ji}} = \exp(-\Delta F_{ij}/k_{\mathrm{B}}T) \tag{23}$$

for transitions induced by an infinite fixed bath at temperature T (ref. 88), with some treatments generalizing to multiple independent reservoirs[11], sometimes of different types. This equation is the core of LDB as it is often used and makes it possible to equate δA_{ij} defined above with the $\delta Q/T$ (refs. 12,89) for transitions induced by an ideal heat bath. The use of the term 'local detailed balance' comes from analogy with true detailed balance, in which a nearly identical expression relates the ratio of transition rates to a Boltzmann factor of energetic difference, arising from full time reversibility of equilibrium dynamics.

Here is a summary of the key ideas in stochastic thermodynamics. Stochastic thermodynamics analyses the time irreversibility of Markov dynamics, derives relationships between flux and irreversibility and connects irreversibility to physical free energy change through the use of LDB and idealized baths (Fig. 3b). Owing to these assumptions, the physical interpretation of stochastic thermodynamics is 'near-equilibrium' (Fig. 3a), in the sense that the baths are assumed to be always at equilibrium, which can constrain the system itself to not stray too far from equilibrium[48].

A standard example in stochastic thermodynamics is of molecular motors such as myosin or kinesin. Molecular motors convert chemical energy from ATP into mechanical movement. The dynamics of a motor is modelled as a sequence of mesostates, in a simple case a three-state cycle ABCA in which mesostate A represents the motor right after the mechanical movement, state B is when the motor is bound with an ATP and state C is right after ATP hydrolysis and before the stepping. When the motor is in these mesostates, microscopic fluctuations of the configuration of the motor and the states of the thermal and chemical reservoirs are assumed to evolve infinitely faster than the evolution of the mesostates, reaching local equilibrium. Under this idealized timescale separation, the dynamics of the motor can be modelled by a Markov process[12]. Irreversibility analysis, as introduced earlier, can then be used to relate the speed of the motor to its irreversibility, and the assumption of LDB further relates these to the free energy consumption. This approach allows for analysis of the trade-off between the speed of the motor and its energy efficiency, for example[83].

LDB involves a timescale separation and a detailed balance approximation on the fast timescale, with large local fluxes between fast-exchanging states, relative to smaller global fluxes (Fig. 4a). Non-balance occurs when global fluxes are relatively bigger (Fig. 4b). Although Markov dynamics on the slow timescale could have large time irreversibility (quantified by equation (22)) and nonlinear responses, nevertheless, LDB is a near-equilibrium assumption insofar as it applies only in the limit of zero net forces and gradients on the fast timescale, supposing equilibrated heat baths to derive equation (23). Further from equilibrium, global flows are large (Fig. 4b), as they are

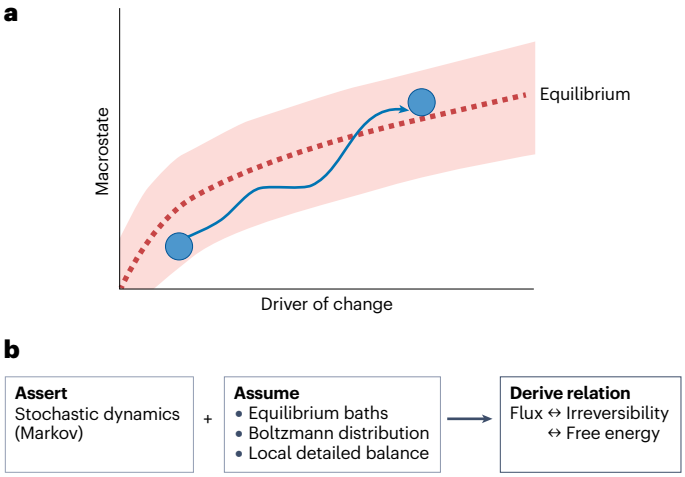

Fig. 3 | Bath-constrained non-equilibrium modelling. a, The orange band indicates dynamical processes that are near equilibrium in the following sense: it is assumed that the endstates and processes have fluctuations that are defined by interactions with ideal thermodynamic reservoirs and obey local detailed balance, to connect the mathematical descriptions with physical quantities. Examples are stochastic thermodynamics and macroscopic fluctuation theory. **b**, Stochastic thermodynamics uses Markov dynamics and local detailed balance to derive relations among flux, irreversibility and locally defined free energy changes. This approach does not generate models; rather, it analyses models and requires equilibrium assumptions to link probabilities to thermodynamic quantities.

in gradient-driven transport described by Fourier's law, Fick's law, Newtonian viscosities or Navier–Stokes hydrodynamics, for example. There are limits to the assumption of a perfect bath (infinitely large and infinitely fast)[48] and to LDB: LDB holds only in the limit of weak driving and/or weak noise[90], LDB may not hold in coarse-grained systems even when there is clear timescale separation[91], and the decomposition of transition rates owing to multiple reservoirs into the sum of transition rates owing to each reservoir is not possible in general[92].

The entropy production rate as a measure of dissipation

The entropy production rate, $\mathrm{d}S_{\mathrm{BG}}(\{p_i(t)\})/\mathrm{d}t = -k_{\mathrm{B}}\,\mathrm{d}D(t)/\mathrm{d}t$ (equation (16)), has historically been considered an important quantity in two ways. First, there was a quest for a non-equilibrium variational principle (reviewed elsewhere[93]); one approach along these lines was to look at $\mathrm{d}S_{\mathrm{BG}}/\mathrm{d}t$ to connect dynamics to dissipation. It turns out that neither the minimization of this entropy production rate[94,95] nor its maximization[96,97] is general as a variational principle (for example, a counterexample is found in bistable chemical systems[98]). Second, entropy dissipation is used to describe irreversibility. But in that regard, $\mathrm{d}S_{\mathrm{BG}}/\mathrm{d}t$ is limited to near equilibrium, which can be seen as follows: equation (22) shows that $D(t)$ gives only the detailed-balance component of irreversibility, not the non-balanced component. Viewed through the lens of experiments, Clausius' entropy is only related to $\delta Q/T$ for reversible equilibrium processes, as in equation (3). Inferencing is only proper and self-consistent for those probability distributions that are maximized, $p_i = p_i^*$, which only holds at equilibrium if the ps represent microstates (as discussed in the section on entropy maximization). A more general and generative principle for non-equilibria is given in the last section, on MaxCal, based on path entropies.

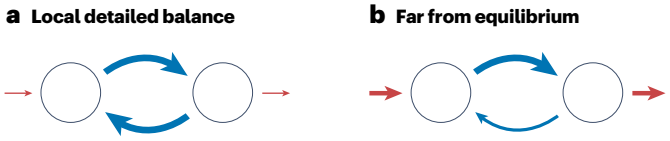

Fig. 4 | Local detailed balance. a, Shown here for a 1D process along the horizontal axis, local detailed balance is the approximation that dynamical exchanges are locally fast (having large symmetric fluxes; blue arrows) relative to more global asymmetric fluxes (red arrows). Local detailed balance is a near-equilibrium approximation as it applies most accurately in the limit as driving forces and gradients approach zero on the microscopic scale. **b**, Far from equilibrium, global flows are larger.

Despite these caveats of limitations to near equilibrium, we emphasize that LDB is only required when there is a need to relate dynamic irreversibility to thermal properties such as free energy change and heat dissipation. Many mathematical results in stochastic thermodynamics, such as equation (19), speed limit relations[58,59], fluctuation relations[14,76,77], uncertainty relations[78,79] and others, remain valid and useful for the analysis of Markov stochastic dynamics in general, even where LDB does not hold.

Jarzynski equality

The Jarzynski equality draws on earlier work relating to computational methods[99,100] and generalizes Clausius' inequality of work W and free energy F into an equality[13]:

$$\langle e^{-W/k_BT} \rangle = e^{-\Delta F/k_BT}. \quad (24)$$

There are technically two versions of the Jarzynski equality (JE): the original (JE_1), as derived in 1997, applies to Hamiltonian dynamics for a system-plus-bath supersystem that starts in equilibrium and gets driven arbitrarily away from equilibrium[14] before relaxing back to equilibrium once the external driving is completed (Fig. 2). This Hamiltonian formulation of the JE_1 does not need the LDB assumptions (although it does need another external perfect heat bath described by temperature T with which the total supersystem can equilibrate), and it can be generalized to strongly coupled systems[101–103]. However, the requirement to start from a Hamiltonian restricts the applicability of this formulation of the JE to colliding particle systems.

The second Jarzynski equality[104] (JE_2) and its corresponding work fluctuation relation[105] are formulated for Markov dynamics, as typically used in stochastic thermodynamics. Because these situations require relating Markov transition probabilities to energetics, JE_2 requires the LDB condition of equation (23). Hence, as with stochastic thermodynamics, JE_2 for Markov dynamics belongs in the classification of Fig. 3, that is, near equilibrium. Other fluctuation relations[106] are similarly restricted.

The paradigmatic example of JE is a rubber band in a heat bath[14]. The rubber band starts at equilibrium with a reservoir and is pulled to a new length, where it is maintained until a new equilibrium is reached. JE dictates how the free energy between the two equilibrium states relates to the work needed to perform this process (which has substantial fluctuation if this 'rubber band' is a macromolecule such as a DNA or protein). JE can infer the free energy difference from specific exponential averages of work or vice versa. The Hamiltonian formulation, JE_1, is valid for any speed of pulling, as long as the system and its heat bath can be considered isolated from the rest of the universe during the process. By contrast, the Markov formulation, JE_2, is only applicable if the pulling is slow enough for the bath to equilibrate at every instance.

Macroscopic fluctuation theory

The field of macroscopic fluctuation theory[15] uses LDT to describe the non-trivial asymptotic behaviours and time irreversibility of stochastic systems in the hydrodynamic limit, in which the particles are so numerous that one can define continuous density and flow. Doing so extends textbook hydrodynamics to the realm of fluctuations of density and flow in physical space-time. Because macroscopic fluctuation theory also assumes a fast equilibrating bath for its physical interpretations, it faces the same LDB limitations as stochastic thermodynamics and therefore can be put in to the same silo as stochastic thermodynamics in Fig. 3. Both stochastic thermodynamics and macroscopic fluctuation theory bring powerful mathematical tools for analysing stochastic processes and their irreversibilities, with results framed in terms of entropy-like quantities that provide direct physical insight in certain circumstances. Although macroscopic fluctuation theory has some variational formulas[15], they are formulated to compute asymptotic statistical properties of density and flow in the hydrodynamic limit, not to generate models of microscopic distributions (discussed subsequently).

Towards a generative statistical physics of non-equilibria

Equilibrium statistical physics has stood alone for more than a century. The rigorous foundations of equilibrium statistical physics, combined with its breadth of applicability and generative power, have, until recently, been unequalled by any treatment for non-equilibrium. This was well articulated by Touchette[16]: 'There is no general principle whereby one can calculate the distribution of the system's states from the sole knowledge of the system's invariants or external constraints imposed on the system .... Such a general principle is precisely what a statistical-mechanical ensemble is, and what is missing from the theory of nonequilibrium systems.'

Said differently, Touchette's aspiration is for a principle that is generative. Such a principle should take as input a model and constraints and generate the probability distribution of routes and rates that are consistent with it. Yes, non-equilibrium physics has specific models — of random flights, Langevin and Fokker–Planck diffusion and others — and useful mathematical results for their analysis, many from stochastic thermodynamics, LDT and macroscopic fluctuation theory. But, as indicated by Figs. 2 and 3, such modelling has largely been limited to starting and ending points of processes that are at or near equilibrium, which require contact with a perfect bath. Such theories do not derive their probability distributions of rates and routes from a model and data, as MaxEnt does for equilibrium. Rather, these approaches assume distributions. To be clear: NEQ modelling can be parsed into two types. The first type, which includes master equations and Markov models, for example, makes no attempt to relate distributions and transition rates to temperature, energies or free energies, there is no requirement for equilibrium heat baths; these dynamics are general. The second type consists of the situations described earlier, of modelling that aims to relate dynamical populations to energies or free energies — in Langevin, Einstein and Smoluchowski diffusion, activated rate theories and stochastic thermodynamics, for example. As described earlier, these approaches assume equilibrium temperature baths, reflected by factors of k_BT appearing in equations derived with these methods.

Touchette's aspiration is for some enveloping principle for making models that can derive non-equilibrium dynamical distributions from the model and data. Furthermore, for broadest generality, a non-equilibrium principle should apply to force–flow relations that go beyond fluids, particles, electrical charges and heat, to encompass also the flows of vehicles and goods along networks, of proteins and genes in cells, of electrical signals in brains and of biological changes in organisms in evolution and ecology. These requirements are fulfilled by MaxCal, as we now describe.

MaxCal is the variational principle for non-equilibria

A general and generative variational method for non-equilibrium is MaxCal (Fig. 5), summarized subsequently. MaxCal[93] does for trajectories what classic MaxEnt does for states. It bears close resemblance to MaxEnt except that the probabilities involved refer to paths through state space, rather than single states; similarly, the constraints often take the form of averages of rate properties instead of state properties.

Here is an example of applying the MaxCal procedure. Consider a two-state, discrete-time process in the dog–flea model[107] in which a single flea (particle) starts at dog a (position along the x-axis) and stochastically jumps between two dogs labelled a and b over N time steps. The observable $\langle N_{ij}\rangle$ is the average number of times in which the flea jumps from dog i to dog j, where the possible transitions are $ij \in \{aa, ab, ba, bb\}$. Implementing these constraints with Lagrange multipliers $\{\beta_{ij}\}$, the caliber is maximized in analogy with minimizing the free energy in MaxEnt:

$$\mathcal{C} = -\sum_{\kappa} p_{\kappa} \ln \frac{p_{\kappa}}{q_{\kappa}} - \beta_{aa} \sum_{\kappa} p_{\kappa} N_{aa\kappa} - \beta_{ab} \sum_{\kappa} p_{\kappa} N_{ab\kappa} - \beta_{ba} \sum_{\kappa} p_{\kappa} N_{ba\kappa} - \beta_{bb} \sum_{\kappa} p_{\kappa} N_{bb\kappa}, \tag{25}$$

in which p_{κ} and q_{κ} are the posterior and prior probability distributions over paths labelled by κ, and $N_{ij\kappa}$ denotes the number of transitions ij in path κ (the normalization constraint is omitted here for the sake of brevity).

Assuming a uniform prior $q_{\kappa} = 1/2^N$, the maximizing path distribution is a Boltzmann-like distribution

$$p^*_{\kappa} = \frac{(e^{-\beta_{aa}})^{N_{aa\kappa}} (e^{-\beta_{ab}})^{N_{ab\kappa}} (e^{-\beta_{ba}})^{N_{ba\kappa}} (e^{-\beta_{bb}})^{N_{bb\kappa}}}{Z_{\mathrm{d}}}, \tag{26}$$

in which

$$Z_{\mathrm{d}} = \sum_{\kappa} (e^{-\beta_{aa}})^{N_{aa\kappa}} (e^{-\beta_{ab}})^{N_{ab\kappa}} (e^{-\beta_{ba}})^{N_{ba\kappa}} (e^{-\beta_{bb}})^{N_{bb\kappa}} \tag{27}$$

is the dynamical partition function. Akin to the relationship between free energy and partition function in MaxEnt, the logarithm of the dynamical partition function is equal to the caliber $\mathcal{C} = \ln Z_{\mathrm{d}}$ (up to a constant $N \ln 2$ from the prior). In further analogy, partial derivatives of the caliber with respect to the Lagrange multipliers give the statistical cumulants of the observables N_{ij}. For example, the first and second partial derivatives give the means and the variances

$$\langle N_{ij}\rangle = \frac{\partial \mathcal{C}}{\partial(-\beta_{ij})} \quad \text{and} \quad \sigma^2_{N_{ij}} = \frac{\partial^2 \mathcal{C}}{\partial(-\beta_{ij})^2}. \tag{28}$$

This equation can be inverted to express β_{ij} as functions of the averages $\langle N_{ij}\rangle$. Then, the posterior path distribution p^*_{κ} of equation (26) is expressed in terms of the observables $\langle N_{ij}\rangle$ and can be plugged back into the path entropy $\mathcal{S} = -\sum_{\kappa} p_{\kappa} \ln (p_{\kappa}/q_{\kappa})$ to obtain the maximized path entropy $\mathcal{S}^*(\langle N_{ij}\rangle)$ as a function of the observables $\langle N_{ij}\rangle$. Partial derivatives show that the multipliers are the path entropic forces:

$$\beta_{ij} = \frac{\partial \mathcal{S}^*}{\partial \langle N_{ij}\rangle}. \tag{29}$$

Maxwell partial derivative relations follow from equations (28) and (29).

Other implications of MaxCal have been reviewed elsewhere[17,18,40,108–110]; here is just a brief summary. First, MaxCal recapitulates – as it should – many known results of near-but-not-at-equilibrium physics[18,108], including the Green–Kubo fluctuation–dissipation relations, Onsager reciprocal relations, Prigogine's minimum entropy-production principle in certain situations and Kirchhoff's current law. Second, because it is generative of models and allows for large forces and gradients, MaxCal can, in principle, derive the phenomenological laws of force–flow relations, such as Newtonian fluid viscosities and Fourier's law of heat flow. For example, Fick's law of particle flows has been derived from a model of independent-particle hopping with inference from MaxCal[107,111], which also gives previously unknown results for the few-particle limit that have been subsequently validated in single-particle experiments[112]. Similarly, other workhorse methods of non-equilibrium statistical physics – such as master equations, Markov models and Fokker–Planck and Langevin equations – can be grounded in models of individual particle trajectories or extended to more complex systems by MaxCal[107,113–115]. MaxCal handles nonlinearities in gene circuits[18,116–118], that is, in toggle switches,

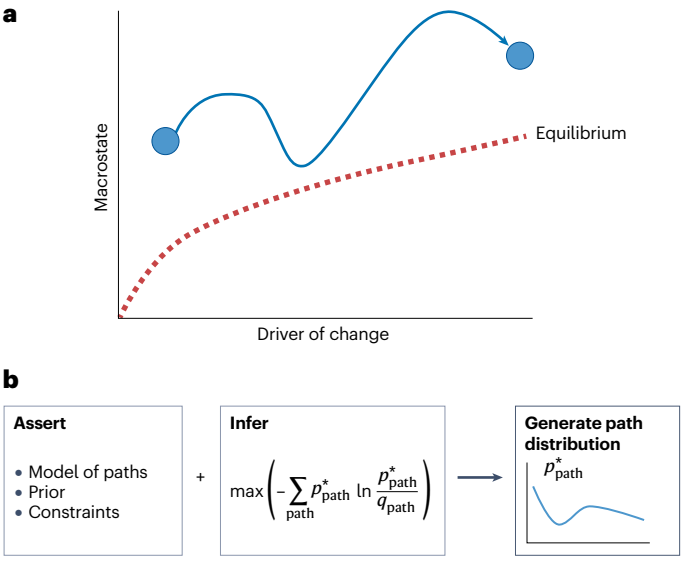

Fig. 5 | The generative inference pipeline of non-equilibrium statistical physics, which generates the pathway distribution from the input assertions. a, The pipeline applies to arbitrary endstates and processes. **b**, The modeller asserts allowed trajectories and accompanying constraints in terms of those trajectories, along with a prior distribution q_{path}; the MaxCal procedure then infers the unique probability distribution p^*_{path} that satisfies the constraints (such as data) and the laws of probability. MaxCal is a general, generative inference principle for non-equilibria, playing the same role as MaxEnt does for equilibria (Fig. 1).

positive and negative feedback situations and Michaelis–Menten binding, as well as nonlinearities in anomalous diffusion and superstatistics[40,119]. It can also give useful insights from molecular simulation trajectories[10,120–122]. Finally, for problems that are currently treated by bath-constrained modelling, assuming a Boltzmann-distributed equilibrium bath, MaxCal can provide next-order corrections[48,123].

Practicalities of formulating models and constraints

As noted earlier, the burdens of modelling with MaxEnt or MaxCal lie with the user. The user defines the model. The user must also assert what process and change are of interest; these inform the choice of constraints. Here, we give some of the relevant considerations.

For equilibrium situations, it is useful to inspect the Boltzmann factor ε_i/k_BT, in which ε_i is the spectrum of allowed energy values for a system. The denominator represents the influence on the system of a heat bath, representing how the environment imposes a fixed average energy. A bath is an assertion about the relevant variables that drive a system. If one considers only the temperature when attempting to model a process in which particle number N is also relevant, one's predictions will err. When N is relevant, one must assert two baths — one holding average energy fixed by k_BT, and the other holding average particle number fixed, having denominator k_BT/μ, in which μ is the chemical potential. This issue is not a flaw of MaxEnt; the user must determine which, and how many, baths are driving the process of interest, or whatever other physics is imposing relevant constraints.

Now consider the numerator in the Boltzmann factor, ϵ_i. These numerators cannot be arbitrary observables; they must satisfy three requirements. The first is conservation: ε_i must represent a conserved quantity, for which 'stuff' can move from one bin to another bin, but cannot appear or disappear. However, conservation alone is not sufficient to guide the choice of constraints. For example, why is energy constrained, rather than momentum? This question is answered by the second requirement, that microscopics must match macroscopics: namely, ε_i must represent a physical property the average of which is exchanged at the boundary of the system. If energy is what is being exchanged at the boundary, then energy must be the ladder of the microscopic model. The third requirement is for mechanistic underpinnings: to give insights about physics or mechanism, the index i must have some meaningful relationship to a model. For example, Maxwell–Boltzmann statistics arise from the kinetic energy of a classical particle of mass m and velocity v, $\varepsilon(v) = mv^2/2$. This relationship enables the insights of statistical physics, explaining macroscopic phenomena in terms of microscopic mechanisms.

Now, for non-equilibrium processes, what are the considerations? Just as in equilibria, where it is necessary to specify constraints, say (T, p, N), so too is it necessary for non-equilibria. Non-equilibrium systems are characterized by average net flow or flows into or out of systems, and such average flows are often part of the relevant constraints for a particular process. This situation contrasts with equilibrium, in which baths hold some average property fixed with no net flow. Or, in non-thermal flows, such as traffic on networks, there is no heat bath at all. Nevertheless, the procedure in MaxCal is often similar to that of MaxEnt, seeking the 'right ensemble' with the appropriate and relevant fluxes. Consider a ball falling through a viscous liquid with average velocity $\langle v \rangle$. Now, if the same ball is dropping through a more viscous liquid, but at the same velocity $\langle v \rangle$ (hence it must be driven by a larger force), the average velocity alone of the ball is not sufficient to characterize the system — the other relevant variable is heat dissipation[124].

The same qualities apply to the choice of constraints in MaxCal as they do for MaxEnt. Conservation: the user seeks to characterize the flows of conserved quantities, such as particles or energy. In the aforementioned dog–flea example, the constraints represent, in a sense, conservation of the flea — it must begin somewhere, and it must end somewhere. Microscopics match macroscopics: for example, in the presence of macroscopic electrical current, the microscopic model should account for possible distributions of electrical current. In the dog–flea example, an ensemble of fleas will be exchanged between the two dogs in some average way: this average is connected to the different microscopic trajectories of individual exchanges. Mechanistic insight: for example, the different trajectories of a molecular motor might have a particular dependence on the ATP concentration driving it. In the dog–flea example, the entropic forces β_{ij} could be connected to the microscopic energy cost of jumps with an energetic model of the flea jumping mechanism.

Outlook

As applications of statistical physics continue to grow, the underlying principles also advance, on multiple fronts. Probabilities have come to be seen through the lens of inference, rather than as frequencies of replicable events. And maximizing the BG entropy is seen as the unique procedure for inference about microscopic distributions that ensure faithfulness to a model, given data and the laws of probability. Assumptions of thermodynamic baths and sometimes of LDB have allowed for generalizations beyond Clausius' slow reversible processes between equilibrium states, but there are still more general approaches. Maximizing entropy is a generative procedure for equilibria, but also for non-equilibria, when taken over pathways (MaxCal), rather than over states (MaxEnt). This approach allows for modelling far from equilibrium, handles nonlinearities of forces and flows, is not bound to baths and functions at the level of individual agents and their trajectories.

We hope that the contents of this Review will foster connections between researchers in different groups within the field of statistical physics. For example, the stochastic thermodynamics community continues to be a source of important and powerful results, but the possible limits of the applicability of those results should be kept squarely in mind. When one approach is lacking, researchers can turn to other methods outlined in this Review — including particularly inferential methods — to fill in the gaps.

On the subject of inferential methods, although there has already been much work described above showing how MaxCal recovers important principles in statistical physics that were already known from previous studies, there remains great untapped potential to bring MaxCal to bear on new theoretical discoveries in non-equilibrium physics. We mentioned the beginnings of a few such results, but further study of MaxCal promises to yield continued theoretical insights in the coming years.

Published online: 1 May 2024

References

1. Huang, K. *Statistical Mechanics* (Wiley, 1987).
2. Hill, T. L. *Statistical Mechanics: Principles and Selected Applications* (Dover Publications, 1987).
3. McQuarrie, D. A. *Statistical Mechanics* (University Science Books, 2000).
4. Dill, K. A. & Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience* 2nd edn (Garland Science, 2010).
5. Pathria, R. K. & Beale, P. D. *Statistical Mechanics* (Academic Press, 2021).
6. Martino, A. D. & Martino, D. D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **4**, e00596 (2018).

7. Harte, J. & Newman, E. A. Maximum information entropy: a foundation for ecological theory. *Trends Ecol. Evol.* **29**, 384–389 (2014).
8. Golan, A. & Harte, J. Information theory: a foundation for complexity science. *Proc. Natl Acad. Sci. USA* **119**, e2119089119 (2022).
9. Fodor, É., Jack, R. L. & Cates, M. E. Irreversibility and biased ensembles in active matter: insights from stochastic thermodynamics. *Annu. Rev. Condens. Matter Phys.* https://doi.org/10.1146/annurev-conmatphys-031720-032419 (2022).
10. Tsai, S.-T., Fields, E., Xu, Y., Kuo, E.-J. & Tiwary, P. Path sampling of recurrent neural networks by incorporating known physics. *Nat. Commun.* **13**, 7231 (2022).
11. Esposito, M. Stochastic thermodynamics under coarse graining. *Phys. Rev. E* **85**, 041125 (2012).
12. Seifert, U. Stochastic thermodynamics: from principles to the cost of precision. *Phys. A Stat. Mech. Appl.* **504**, 176–191 (2018).
13. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
14. Jarzynski, C. Equalities and inequalities: irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.* **2**, 329–351 (2011).
15. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G. & Landim, C. Macroscopic fluctuation theory. *Rev. Mod. Phys.* **87**, 593–636 (2015).
16. Touchette, H. The large deviation approach to statistical mechanics. *Phys. Rep.* **478**, 1–69 (2009).
17. Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).
18. Ghosh, K., Dixit, P. D., Agozzino, L. & Dill, K. A. The maximum caliber variational principle for nonequilibria. *Annu. Rev. Phys. Chem.* **71**, 213–238 (2020).
19. Bagchi, B. *Statistical Mechanics for Chemistry and Materials Science* (CRC Press, 2018).
20. Reichl, L. *A Modern Course in Statistical Physics* (Wiley, 2016).
21. Kubo, R. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn* **12**, 570–586 (1957).
22. Yang, Y.-J. & Qian, H. Statistical thermodynamics and data ad infinitum: conjugate variables as forces, and their statistical variations. Preprint at http://arxiv.org/abs/2205.09321 (2023).
23. Brush, S. G. *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the Nineteenth Century, Book 1: Physics and the Atomists* (North Holland, 1986).
24. Kolmogorov, A. N. *Foundations of the Theory of Probability: Second English Edition* (Dover Publications, 2018).
25. Inaba, H. The development of ensemble theory. *Eur. Phys. J. H* **40**, 489–526 (2015).
26. Lu, Z. & Qian, H. Emergence and breaking of duality symmetry in generalized fundamental thermodynamic relations. *Phys. Rev. Lett.* **128**, 150603 (2022).
27. Sanov, I. N. On the probability of large deviations of random variables. *Mat. Sbornik* **42**, 11–44 (1958).
28. Landau, L. D. & Lifshitz, E. M. *Statistical Physics* 3rd edn, Part 1: Vol. 5 (Butterworth-Heinemann, 1980).
29. Csiszar, I., Cover, T. & Choi, B. S. Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory* **33**, 788–801 (1987).
30. Chetrite, R. & Touchette, H. Nonequilibrium microcanonical and canonical ensembles and their equivalence. *Phys. Rev. Lett.* **111**, 120601 (2013).
31. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
32. Denbigh, K. & Denbigh, J. *Entropy in Relation to Incomplete Knowledge* (Cambridge Univ. Press, 1985).
33. Shore, J. & Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theor.* **26**, 26–37 (1980).
34. Bak, P. *How Nature Works: The Science of Self-organized Criticality* (Copernicus, 1996).
35. Tsallis, C. Nonextensive statistical mechanics: construction and physical interpretation. In *Nonextensive Entropy: Interdisciplinary Applications* (Oxford Univ. Press, 2004).
36. Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Nonadditive entropies yield probability distributions with biases not warranted by the data. *Phys. Rev. Lett.* **111**, 180604 (2013).
37. Tsallis, C. Conceptual inadequacy of the Shore and Johnson axioms for wide classes of complex systems. *Entropy* **17**, 2853–2861 (2015).
38. Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Reply to C. Tsallis' 'Conceptual inadequacy of the Shore and Johnson axioms for wide classes of complex systems'. *Entropy* **17**, 5043–5046 (2015).
39. Jizba, P. & Korbel, J. Maximum entropy principle in statistical inference: case for non-Shannonian entropies. *Phys. Rev. Lett.* **122**, 120601 (2019).
40. Caticha, A. Entropy, information, and the updating of probabilities. *Entropy* **23**, 895 (2021).
41. Peterson, J., Dixit, P. D. & Dill, K. A. A maximum entropy framework for nonexponential distributions. *Proc. Natl Acad. Sci. USA* **110**, 20380–20385 (2013).
42. Van der Straeten, E. & Beck, C. Superstatistical distributions from a maximum entropy principle. *Phys. Rev. E* **78**, 051101 (2008).
43. Davis, S. Conditional maximum entropy and superstatistics. *J. Phys. A Math. Theor.* **53**, 445006 (2020).
44. Davis, S. & Gutiérrez, G. Temperature is not an observable in superstatistics. *Phys. A Stat. Mech. Appl.* **505**, 864–870 (2018).
45. Davis, S. The *q*-canonical ensemble as a consequence of Bayesian superstatistics. Preprint at http://arxiv.org/abs/2112.11496 (2022).
46. Ramshaw, J. D. Maximum entropy and constraints in composite systems. *Phys. Rev. E* **105**, 024138 (2022).
47. Vasconcelos, G. L., Salazar, D. S. P. & Macêdo, A. M. S. Maximum entropy approach to *H*-theory: statistical mechanics of hierarchical systems. *Phys. Rev. E* **97**, 022104 (2018).
48. Pachter, J. A. & Dill, K. A. Nonequilibrium statistical physics beyond the ideal heat bath approximation. *Phys. Rev. E* **107**, 014131 (2023).
49. Umpierrez, H. & Davis, S. Fluctuation theorems in q-canonical ensembles. *Phys. A Stat. Mech. Appl.* **563**, 125337 (2021).
50. Bercher, J. F. Tsallis distribution as a standard maximum entropy solution with 'tail' constraint. *Phys. Lett. A* **372**, 5657–5659 (2008).
51. Hernando, A. & Plastino, A. Variational principle underlying scale invariant social systems. *Eur. Phys. J. B* **85**, 293 (2012).
52. Davis, S. & Gutiérrez, G. Emergence of Tsallis statistics as a consequence of invariance. *Phys. A Stat. Mech. Appl.* **533**, 122031 (2019).
53. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 2nd edn (Wiley-Interscience, 2006).
54. Schnakenberg, J. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.* **48**, 571–585 (1976).
55. Parrondo, J. M. R., Horowitz, J. M. & Sagawa, T. Thermodynamics of information. *Nat. Phys.* **11**, 131–139 (2015).
56. Amari, S.-I. *Information Geometry and Its Applications*. *Applied Mathematical Sciences* (Springer Japan, 2016).
57. Crooks, G. E. Measuring thermodynamic length. *Phys. Rev. Lett.* **99**, 100602 (2007).
58. Nicholson, S. B., García-Pintos, L. P., del Campo, A. & Green, J. R. Time-information uncertainty relations in thermodynamics. *Nat. Phys.* **16**, 1211–1215 (2020).
59. Ito, S. & Dechant, A. Stochastic time evolution, information geometry, and the Cramer–Rao bound. *Phys. Rev. X* **10**, 021056 (2020).
60. Freidlin, M. I. & Wentzell, A. D. *Random Perturbations of Dynamical Systems* (Springer-Verlag, 1984).
61. Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).
62. Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer-Verlag, 2009).
63. Nolting, B. C. & Abbott, K. C. Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology* https://doi.org/10.1890/15-1047.1 (2015).
64. Fang, X., Kruse, K., Lu, T. & Wang, J. Nonequilibrium physics in biology. *Rev. Mod. Phys.* **91**, 045004 (2019).
65. Agozzino, L., Balázsi, G., Wang, J. & Dill, K. A. How do cells adapt? Stories told in landscapes. *Annu. Rev. Chem. Biomol. Eng.* **11**, 155–182 (2020).
66. Qian, H., Cheng, Y.-C. & Yang, Y.-J. Kinematic basis of emergent energetics of complex dynamics. *EPL* **131**, 50002 (2020).
67. Yang, Y.-J. & Cheng, Y.-C. Potentials of continuous Markov processes and random perturbations. *J. Phys. A Math. Theor.* **54**, 195001 (2021).
68. Seifert, U. From stochastic thermodynamics to thermodynamic inference. *Annu. Rev. Condens. Matter Phys.* **10**, 171–192 (2019).
69. Peliti, L. & Pigolotti, S. *Stochastic Thermodynamics: An Introduction* (Princeton Univ. Press, 2021).
70. Prigogine, I. & Defay, R. *Chemical Thermodynamics* (Wiley, 1962).
71. Jiang, D.-Q., Qian, M. & Qian, M.-P. *Mathematical Theory of Nonequilibrium Steady States: On the Frontier of Probability and Dynamical Systems* (Springer, 2004).
72. Yang, Y.-J. & Qian, H. Bivectorial nonequilibrium thermodynamics: cycle affinity, vorticity potential, and Onsager's principle. *J. Stat. Phys.* **182**, 46 (2021).
73. Wang, J., Xu, L. & Wang, E. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl Acad. Sci. USA* **105**, 12271–12276 (2008).
74. Ge, H. Extended forms of the second law for general time-dependent stochastic processes. *Phys. Rev. E* **80**, 021137 (2009).
75. Esposito, M. & Van den Broeck, C. Three detailed fluctuation theorems. *Phys. Rev. Lett.* https://doi.org/10.1103/PhysRevLett.104.090601 (2010).
76. Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **61**, 2361–2366 (2000).
77. Yang, Y.-J. & Qian, H. Unified formalism for entropy production and fluctuation relations. *Phys. Rev. E* **101**, 022129 (2020).
78. Barato, A. C. & Seifert, U. Thermodynamic uncertainty relation for biomolecular processes. *Phys. Rev. Lett.* **114**, 158101 (2015).
79. Horowitz, J. M. & Gingrich, T. R. Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nat. Phys.* **16**, 15–20 (2020).
80. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl Acad. Sci. USA* **71**, 4135–4139 (1974).
81. Sartori, P. & Pigolotti, S. Thermodynamics of error correction. *Phys. Rev. X* **5**, 041039 (2015).
82. Lan, G., Sartori, P., Neumann, S., Sourjik, V. & Tu, Y. The energy-speed-accuracy tradeoff in sensory adaptation. *Nat. Phys.* **8**, 422–428 (2012).
83. Wagoner, J. A. & Dill, K. A. Mechanisms for achieving high speed and efficiency in biomolecular machines. *Proc. Natl Acad. Sci. USA* **116**, 5902–5907 (2019).
84. Brown, A. I. & Sivak, D. A. Theory of nonequilibrium free energy transduction by molecular machines. *Chem. Rev.* **120**, 434–459 (2020).
85. Yang, Y.-J. & Qian, H. Time-translational symmetry in statistical dynamics dictates Einstein relation, Green–Kubo formula, and their generalizations. *Phys. Rev. E* **107**, 024110 (2023).

86. Graham, R. Covariant stochastic calculus in the sense of Itô. *Phys. Lett. A* **109**, 209–212 (1985).
87. Ao, P. Potential in stochastic differential equations: novel construction. *J. Phys. A Math. Gen.* **37**, L25–L30 (2004).
88. Cockrell, C. & Ford, I. J. Stochastic thermodynamics in a non-Markovian dynamical system. *Phys. Rev. E* **105**, 064124 (2022).
89. Sekimoto, K. *Stochastic Energetics* (Springer, 2010).
90. Falasco, G. & Esposito, M. Local detailed balance across scales: from diffusions to jump processes and beyond. *Phys. Rev. E* **103**, 042114 (2021).
91. Hartich, D. & Godec, A. Violation of local detailed balance upon lumping despite a clear timescale separation. *Phys. Rev. Res.* **5**, L032017 (2023).
92. Wasnik, V. Revisiting multiple thermal reservoir stochastic thermodynamics. Preprint at http://arxiv.org/abs/2303.14949 (2023).
93. Jaynes, E. T. The minimum entropy production principle. *Annu. Rev. Phys. Chem.* **31**, 579–601 (1980).
94. Klein, M. J. & Meijer, P. H. E. Principle of minimum entropy production. *Phys. Rev.* **96**, 250–255 (1954).
95. de Groot, S. R. & Mazur, P. *Non-Equilibrium Thermodynamics* (Dover Publications, 2011).
96. Dewar, R. C. Maximum entropy production as an inference algorithm that translates physical assumptions into macroscopic predictions: don't shoot the messenger. *Entropy* **11**, 931–944 (2009).
97. Martyushev, L. M. & Seleznev, V. D. The restrictions of the maximum entropy production principle. *Phys. A Stat. Mech. Appl.* **410**, 17–21 (2014).
98. Vellela, M. & Qian, H. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *J. R. Soc. Interface* **6**, 925–940 (2008).
99. Reinhardt, W. P. & Hunter, J. E. III. Variational path optimization and upper and lower bounds to free energy changes via finite time minimization of external work. *J. Chem. Phys.* **97**, 1599–1601 (1992).
100. Hunter, J. E. III, Reinhardt, W. P. & Davis, T. F. A finite-time variational method for determining optimal paths and obtaining bounds on free energy changes from computer simulations. *J. Chem. Phys.* **99**, 6856–6864 (1993).
101. Seifert, U. First and second law of thermodynamics at strong coupling. *Phys. Rev. Lett.* **116**, 020601 (2016).
102. Jarzynski, C. Stochastic and macroscopic thermodynamics of strongly coupled systems. *Phys. Rev. X* **7**, 011008 (2017).
103. Miller, H. J. D. & Anders, J. Entropy production and time asymmetry in the presence of strong interactions. *Phys. Rev. E* **95**, 062123 (2017).
104. Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **90**, 1481–1487 (1998).
105. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **60**, 2721–2726 (1999).
106. Spinney, R. & Ford, I. Fluctuation relations: a pedagogical overview. In *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* 3–56 (John Wiley & Sons, 2013).
107. Stock, G., Ghosh, K. & Dill, K. A. Maximum caliber: a variational approach applied to two-state dynamics. *J. Chem. Phys.* **128**, 194102 (2008).
108. Hazoglou, M. J., Walther, V., Dixit, P. D. & Dill, K. A. Communication: maximum caliber is a general variational principle for nonequilibrium statistical mechanics. *J. Chem. Phys.* **143**, 051104 (2015).
109. Davis, S., González, D. & Gutiérrez, G. Probabilistic inference for dynamical systems. *Entropy* **20**, 696 (2018).
110. Tapia, I., Gutiérrez, G. & Davis, S. Prediction and retrodiction in statistical mechanics from the principle of maximum caliber. Preprint at http://arxiv.org/abs/2303.09672 (2023).
111. Ghosh, K., Dill, K. A., Inamdar, M. M., Seitaridou, E. & Phillips, R. Teaching the principles of statistical dynamics. *Am. J. Phys.* **74**, 123–133 (2006).
112. Wu, D. et al. Trajectory approach to two-state kinetics of single particles on sculpted energy landscapes. *Phys. Rev. Lett.* **103**, 050603 (2009).
113. Ge, H., Pressé, S., Ghosh, K. & Dill, K. A. Markov processes follow from the principle of maximum caliber. *J. Chem. Phys.* **136**, 064108 (2012).
114. Lee, J. & Pressé, S. A derivation of the master equation from path entropy maximization. *J. Chem. Phys.* **137**, 074103 (2012).
115. González, D., Davis, S. & Gutiírrez, G. Newtonian dynamics from the principle of maximum caliber. *Found. Phys.* **44**, 923–931 (2014).
116. Firman, T., Balázsi, G. & Ghosh, K. Building predictive models of genetic circuits using the principle of maximum caliber. *Biophys. J.* **113**, 2121–2130 (2017).
117. Firman, T., Wedekind, S., McMorrow, T. J. & Ghosh, K. Maximum caliber can characterize genetic switches with multiple hidden species. *J. Phys. Chem. B* **122**, 5666–5677 (2018).
118. Pressé, S., Ghosh, K. & Dill, K. A. Modeling stochastic dynamics in biochemical systems with feedback using maximum caliber. *J. Phys. Chem. B* **115**, 6202–6212 (2011).
119. Sattin, F. Bayesian approach to superstatistics. *Eur. Phys. J. B Condens. Matter Complex Syst.* **49**, 219–224 (2006).
120. Bolhuis, P. G., Brotzakis, Z. F. & Vendruscolo, M. A maximum caliber approach for continuum path ensembles. *Eur. Phys. J. B* **94**, 188 (2021).
121. Alves, A., Dias, A. G. & da Silva, R. Maximum entropy principle and the Higgs boson mass. *Phys. A Stat. Mech. Appl.* **420**, 1–7 (2015).
122. Tiwary, P. & Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl Acad. Sci. USA* **113**, 2839–2844 (2016).
123. Dixit, P. D. A maximum entropy thermodynamics of small systems. *J. Chem. Phys.* **138**, 184111 (2013).
124. Agozzino, L. & Dill, K. Minimal constraints for maximum caliber analysis of dissipative steady-state systems. *Phys. Rev. E* **100**, 010105 (2019).

Acknowledgements
The authors are grateful to the Stony Brook University Laufer Center for support. The authors also thank B. Cannon, C. Jarzynski, C. Kocher, S. Pressé, D. Sivak and J. Wang for insightful comments and helpful feedback.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Physics* thanks Peter Bolhuis, Gonzalo Gutierrez Gallardo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.