Statistical Physics of Adaptation

Nikolay Perunov, Robert A. Marsland, and Jeremy L. England

Department of Physics, Physics of Living Systems Group, Massachusetts Institute of Technology, Floor 6,
400 Tech Square, Cambridge, Massachusetts 02139, USA

Whether by virtue of being prepared in a slowly relaxing, high-free energy initial condition, or because they are constantly dissipating energy absorbed from a strong external drive, many systems subject to thermal fluctuations are not expected to behave in the way they would at thermal equilibrium. Rather, the probability of finding such a system in a given microscopic arrangement may deviate strongly from the Boltzmann distribution, raising the question of whether thermodynamics still has anything to tell us about which arrangements are the most likely to be observed. In this work, we build on past results governing nonequilibrium thermodynamics and define a generalized Helmholtz free energy that exactly delineates the various factors that quantitatively contribute to the relative probabilities of different outcomes in far-from-equilibrium stochastic dynamics. By applying this expression to the analysis of two examples—namely, a particle hopping in an oscillating energy landscape and a population composed of two types of exponentially growing self-replicators—we illustrate a simple relationship between outcome-likelihood and dissipative history. In closing, we discuss the possible relevance of such a thermodynamic principle for our understanding of self-organization in complex systems, paying particular attention to a possible analogy to the way evolutionary adaptations emerge in living things.

Subject Areas: Biological Physics, Complex Systems,
Statistical Physics

I. INTRODUCTION

It has long been understood that nonequilibrium driving can bring about novel patterns of complex organization in thermally fluctuating many-body systems. Whether in fluid dynamics [1,2] or so-called "active matter" mixtures [3,4], numerous examples have been characterized where qualitatively new types of collective behavior become possible because a large set of interacting degrees of freedom are allowed to act together as a conduit for chemical and heat dissipation in surrounding reservoirs.

Many have hoped that statistical thermodynamics might provide us with general predictive principles for the behavior of such systems, much in the same way as in the equilibrium case where the Boltzmann distribution applies. However, historically, efforts along these lines have encountered serious difficulty outside of the near-equilibrium, linear-response regime [5] because the number of time, length, and energy scales that are important in determining the dynamics can multiply confoundingly once external drives become arbitrarily strong.

Nonetheless, there is renewed reason to anticipate that a thermodynamic universality might yet be recognized in strongly driven systems that exhibit complex collective behavior. In the last two decades, many fundamental results in nonequilibrium statistical mechanics have been derived [6–15], for the first time establishing general relationships between far-from-equilibrium probability distributions over states and the thermodynamic fluxes that accompany their dynamical evolution. The common theme in these findings is that time-reversal symmetry and energy conservation together imply that the rates of stochastic transitions in driven open systems have quantitative relationships with amounts of energy and matter exchanged with surrounding baths. In particular, it has been shown that transitions that are more statistically irreversible must be accompanied by more positive amounts of total entropy production.

We aim here to suggest a new interpretation for one of these results that has potential relevance to our understanding of the organization of driven many-body systems. In particular, we point out that the likelihood of observing a given structure to emerge in nonequilibrium evolution is strongly influenced by the amount of absorption and dissipation of work during its history of formation. We examine the mechanism of this general relationship in simple analytical examples. Subsequently, taking inspiration from the way evolutionary adaptation is understood in a biological context, we argue that many structures formed far from equilibrium may appear to have been specially selected for physical properties connected to their ability to absorb work from the particular driving environment, and we discuss the relevance of this hypothesis to studying the physics of self-organization.

II. DISSIPATION AND DRIFT IN A TIME-VARYING FIELD

At thermal equilibrium, it is well known that the likelihood of observing a system in one of its microstates has a simple exponential relationship to the energy of that microstate. Once external drives push the system away from equilibrium, however, the relationship between probability and the flow of energy in the system, in general, becomes more complicated. In order to start to see how, let us first consider a simple model of a single particle stochastically hopping in a landscape of discrete energy levels. Suppose, for such a system, that state x_i has a particular energy E_i, and each pair of states x_i and x_j are separated by an activation barrier of energy $B_{ij} \equiv B_{ji} > \max(E_i, E_j)$. To couple to a thermal bath of temperature $T \equiv 1/\beta$ in a way that is consistent with our underlying assumptions about time-reversibility, we make the Arrhenius assumption that, when our particle is in state x_i, it always has a constant probability rate $r_{i \to j}$ of stochastically hopping to state x_j, which is given by [16]

$$r_{i \to j} = r_{ij}^0 \exp\left[-\beta(B_{ij} - E_i)\right]. \tag{1}$$

Here, $r_{ij}^0 = r_{ji}^0$ is a rate constant specific to the pair of states that remains unaffected when we change E_i, E_j, or B_{ij}. By assuming transition probability rates to take this form, we are guaranteed that, so long as the energies of states and barriers do not change with time, the probability distribution for the location of our hopping particle must, at steady state, converge on a Boltzmann distribution $p_s(x_i) \propto \exp[-\beta E_i]$ that will obey the detailed balance condition $p_s(x_i)/p_s(x_j) = r_{j \to i}/r_{i \to j}$ required by time-reversal symmetry.

To drive the system away from equilibrium, we choose the energies $E_i(t)$ and $B_{ij}(t)$ to be arbitrary functions of time. Work is done on the system whenever the energy E_i changes while the particle is located at x_i, and heat $\Delta Q = E_i(t) - E_j(t)$ is exhausted into the surrounding thermal bath whenever the particle hops from x_i to x_j at time t. In particular, we consider an energy landscape composed of only three states x_1, x_2, x_3 arranged in a row so that $r_{12}^0 = r_{23}^0 = r > 0$ and $r_{13}^0 = 0$ (Fig. 1). All three states are initially assumed to have the same energy $E_i = 0$ and to be interlaced with two barriers of equal height $B_{12} = B_{23} = \Delta E > 0$.

If the single particle is placed at x_2 at $t = 0$, then, in the absence of any external drives, it is obviously equally likely to hop to x_1 on "the left" or to x_3 on "the right" in some short time $\tau \ll 1/r$ since $r_{2 \to 1} = r_{2 \to 3} = re^{-\beta \Delta E}$. Moreover, the probability of returning to x_2 in time τ after making that first jump is also the same for a particle at x_1 or x_3. We may then choose to drive the system at high frequency so that $E_1(t) = -\Delta E \cos(\omega t)/2$ and $B_{12}(t) = \Delta E - \Delta E \cos(\omega t)/2$, while E_2, E_3, and B_{23} remain constant in time and $\omega \gg 1/\tau \gg r$.
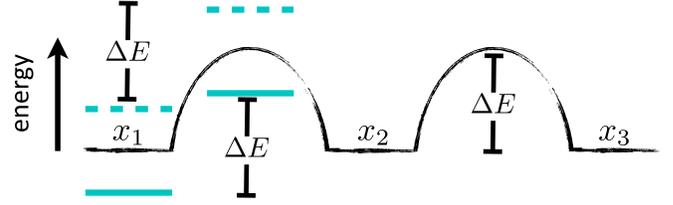

FIG. 1. Three separate microstates initially of equal energy $E = 0$ are arranged on a line so that stochastic transitions may take place between adjacent locations separated by activation barriers of height ΔE. When state x_1 and the barrier separating it from x_2 are driven, such that their energies oscillate in time, a preference for x_1 over x_3 as a destination for finite time evolution from x_2 develops. This statistical drift is necessarily accompanied by the conversion of absorbed work into dissipated heat in the surroundings. The energies of the driven states are drawn at their extremal values in blue, with dashed lines and solid lines, respectively, in phase with each other.

The simple point here is that by contriving the external drive in the manner described, we have necessarily also introduced a strong correlation between dissipation and drift. To see this, we note that the rightward hopping rate $r_{2 \to 3} = re^{-\beta \Delta E}$ does not change with time. In contrast, the leftward hopping rate gets an extra boost from the oscillation of the barrier energy $B_{12}(t)$, so that

$$r_{2 \to 1}(t) = r \exp\left[-\beta(\Delta E - \Delta E \cos[\omega t]/2)\right]. \tag{2}$$

So long as $\beta \Delta E \gg 1$, leftward hopping events are much more likely to happen during the part of the drive cycle when the hopping rate is maximal because of the exponential rate enhancement that comes from the downswing of $B_{12}(t)$. At this maximum, the ratio of the leftward and rightward rates is

$$\frac{r_{2 \to 1}^{\max}}{r_{2 \to 3}} = \exp\left[\beta \Delta E/2\right], \tag{3}$$

indicating a strong bias towards landing at x_1 rather than at x_3 after initially being placed at x_2. One may confirm this bias by averaging each hopping rate over a whole drive period $2\pi/\omega$, obtaining $\overline{r_{2 \to 1}} = re^{-\beta \Delta E} I_0(\beta \Delta E/2) > \overline{r_{2 \to 3}} = re^{-\beta \Delta E}$.

At the same time, it is clear this bias towards one outcome over the other usually requires extra dissipation that accompanies motion in the likely direction. During moments in the drive cycle when transits from x_2 to x_1 are likely (such as at $t = 0$), our chosen drive has guaranteed that E_1 will be at its minimum value, so a hop to the left must typically be accompanied by a positive amount of dissipative external entropy production $\beta \langle \Delta Q \rangle_{2 \to 1} = \beta \Delta E/2$. However, hops to the right will not dissipate at all since $E_3 = E_2 = 0$ do not change with time, so $\beta \langle \Delta Q_{2 \to 3} \rangle = 0$. Thus, we can see that, at least in this case, consistently positive entropy production in the surroundings and concerted drift towards a more likely

outcome of stochastic evolution are two sides of the same coin.

The notion that the flow of probability density in driven systems goes in directions that lead to more dissipation of heat has some intuitive appeal, but, not surprisingly, it turns out to be too simple to be true in general. In the next section, we present a relationship between probability flow and thermodynamic flux that *does* apply in a very broad class of driven systems, thus illustrating, more generally, what role dissipative history plays in determining the outcome of a nonequilibrium dynamical evolution.

III. ENTROPY PRODUCTION AND STOCHASTIC EVOLUTION

The general scenario of interest to us is one where some large number of interacting particles are confined to a volume V that is held in contact with a heat bath at constant inverse temperature $\beta \equiv 1/T$. In accordance with the approach pioneered in previous works [6], the classical Hamiltonian of the whole setup is written as

$$H_{\text{tot}} = H_{\text{sys}}(\mathbf{x}, \lambda(t)) + H_{\text{bath}}(\mathbf{y}) + h_{\text{int}}(\mathbf{x}, \mathbf{y}), \quad (4)$$

where $\mathbf{x} = \{q_1^{(1)}, p_1^{(1)}, ..., q_N^{(3)}, p_N^{(3)}\}$ accounts for all the coordinate degrees of freedom for the N particles in the system, $\mathbf{y}$ does the same for the bath, and the Hamiltonian functions H_{sys}, H_{bath}, and h_{int} define conservative interactions among the various position coordinates of system and bath. The function $\lambda(t)$ plays the role of a time-varying external field that acts exclusively on the *system* and can do work on the coordinates $\mathbf{x}$. Crucially, h_{int} is assumed to be small, so the term merely plays the formal role of permitting the flow of energy between the bath and the system and may otherwise be ignored.

The standard approach [7] to modeling such a system is to posit that the coupling to the heat bath introduces stochasticity into the observed dynamics, so if one starts off at some particular point in the system's phase space $\mathbf{x}(0)$ at time $t = 0$, then, for a given choice of the driving field $\lambda(t)$, there is some probability density for microtrajectories of the system $\pi_\tau[\mathbf{x}(t)|\mathbf{x}(0); \lambda(t)]$ that expresses how likely one would be to observe the system progressing through a given series of subsequent arrangements $\mathbf{x}(t)$ over time τ. This density is governed by the microscopic reversibility relation put forward by Crooks [7]:

$$\frac{\pi_\tau[\mathbf{x}_\tau^\dagger(\tau - t)|\mathbf{x}^\dagger(\tau); \lambda(\tau - t)]}{\pi_\tau[\mathbf{x}(t)|\mathbf{x}(0); \lambda(t)]}$$

$$= \exp\left(-\beta \Delta Q[\mathbf{x}(t)]\right)$$

$$= \exp\left(-\Delta S_{\text{bath}}[\mathbf{x}(t)]\right). \quad (5)$$

The above equation relates the probability of observing a given microtrajectory to the probability of observing the time-reversed movie trajectory after flipping all momentum coordinates in the starting state $[\mathbf{x}^\dagger(\tau) = (q_1^{(1)}(\tau),$

$-p_1^{(1)}(\tau), ...)]$. Because of the underlying time-reversal symmetry of Newton's laws, the ratio of the forward and reverse probabilities must be exactly equal to the exponential of the heat $\beta \Delta Q[\mathbf{x}(t)]$ evolved into the bath in the forward direction as the system traverses the microtrajectory $\mathbf{x}(t)$. Since this heat is being transferred to a vast external reservoir at constant temperature $T = 1/\beta$, we may accordingly identify it thermodynamically as entropy production and write $\beta \Delta Q[\mathbf{x}(t)] = \Delta S_{\text{bath}}[\mathbf{x}(t)]$. Thus, the Crooks result establishes a general, exact, microscopically detailed relationship between statistical irreversibility and entropy production in the surrounding bath, and for the remainder of this work, we treat heat evolution and external entropy production interchangeably.

Recently [14], we have demonstrated the consequences of the above relation for the stochastic dynamics of arbitrarily driven nonequilibrium macrostates. We may define a macrostate by labeling an arbitrary collection of microstates as sharing some macro-observable property **I**, and in that case, if the system is prepared according to some controlled experimental procedure to be in **I**, then it will implicitly possess some nonequilibrium density over microstates $p(\mathbf{x}|\mathbf{I})d\mathbf{x} \equiv p_i(\mathbf{x})d\mathbf{x}$. If we furthermore designate some other disjoint set of microstates to have macroproperty **II**, then we can define $\pi_\tau[\mathbf{I} \to \mathbf{II}; \lambda(t)]$ as the probability that the system is observed to have property **II** after stochastically evolving under the applied field $\lambda(t)$ for time τ. And if the system is indeed observed to be in **II**, then it will implicitly be distributed over the microstates available to it according to some new density $p(\mathbf{x}|\mathbf{II}, \mathbf{I}; \lambda(t); \tau) \equiv p_f(\mathbf{x})d\mathbf{x}$. In this case, we can also define the probability $\pi_\tau^!\left[\mathbf{II}^\dagger \to \mathbf{I}^\dagger; \lambda(\tau - t)\right]$ of reverting back to **I** if particle momenta are reversed and the applied field is run backward (Fig. 2). With these definitions in hand, it may be shown that
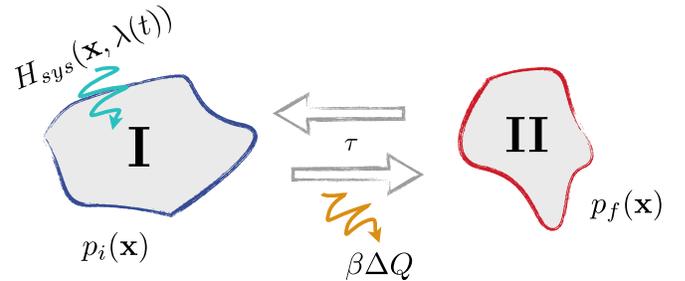

FIG. 2. An arbitrary nonequilibrium macrostate for the system of interest is constructed as the probability density $p_i(\mathbf{x})d\mathbf{x}$ over the microstates $\mathbf{x}$ when the system is prepared by a controlled procedure to have some macroscopic property **I**. The system is externally driven for a period of duration τ by a time-varying field $H_{\text{sys}}(\mathbf{x}, \lambda(t))$ while in contact with a thermal bath of inverse temperature β that absorbs heat ΔQ. Afterwards, one may ask whether the system now has a new macroscopic property **II**; if so, the new probability distribution over microstates is implicitly given by $p_f(\mathbf{x})d\mathbf{x}$.

$$\frac{\pi_\tau^{\mathbf{I}}[\mathbf{II}^\dagger \to \mathbf{I}^\dagger; \lambda(\tau - t)]}{\pi_\tau[\mathbf{I} \to \mathbf{II}; \lambda(t)]} = \langle \exp(-\Delta S_{\text{tot}}) \rangle_{\mathbf{I} \to \mathbf{II}}, \quad (6)$$

where $\Delta S_{\text{tot}}[\mathbf{x}(t)] \equiv \{\ln[p_i(\mathbf{x}(0))]/[p_f(\mathbf{x}(\tau))]\} + \beta \Delta Q[\mathbf{x}(t)]$. Put succinctly, this result says that statistical irreversibility measured on arbitrary macro-observables still has an exact quantitative relationship to entropy production, with two important differences: first, that one must account for the internal entropy difference between macrostates, and second, that now the entropy must be exponentially averaged in the appropriate way over the statistical weights of all the micro-trajectories propagating from the starting macrostate to the ending macrostate. Of course, the fundamental physics behind this statement remains the same as that established by the microscopic relation in Eq. (5). The importance of layering a coarse-graining procedure on top derives only from the fact that experiments on many-body systems invariably involve observing or measuring a limited set of macroscopic criteria or quantities in the system, and so we need to do proper book-keeping on the number of microscopic ways that the system could appear the same to an experimenter at the macro level.

Previously, the above result was used to derive an extension of the second law of thermodynamics, which was applied to studying the thermodynamic constraints obeyed by self-replicators as they grow. Here, we instead use the equation to give a thermodynamic account of driven stochastic evolution. Thus, let us consider a many-particle system prepared in **I** in the manner described above, and let us suppose that we have defined two distinct possible macroscopic outcomes, **II** and **III**, that could lie a long way in the system's future (Fig. 3). A question of great interest to us would be which of the two outcomes should be more likely given $H_{\text{sys}}(\mathbf{x}, \lambda(t))$, that is, given the way the particles inside the system interact with each other and the way the system is being externally driven. Formally, we may write

$$\ln\left[\frac{\pi_\tau[\mathbf{I} \to \mathbf{II}; \lambda(t)]}{\pi_\tau[\mathbf{I} \to \mathbf{III}; \lambda(t)]}\right] = \ln\left[\frac{\pi_\tau^{\mathbf{I}}[\mathbf{II}^\dagger \to \mathbf{I}^\dagger; \lambda(\tau - t)]}{\pi_\tau^{\mathbf{I}}[\mathbf{III}^\dagger \to \mathbf{I}^\dagger; \lambda(\tau - t)]}\right]$$
$$- \ln\left[\frac{\langle \exp(-\Delta S_{\text{tot}}) \rangle_{\mathbf{I} \to \mathbf{II}}}{\langle \exp(-\Delta S_{\text{tot}}) \rangle_{\mathbf{I} \to \mathbf{III}}}\right]. \quad (7)$$

Much can already be learned from the setup of this equation, but a few additional simplifying assumptions help make its structure even clearer: If we assume that the system is driven time symmetrically for a long time, we can dispense with using † operators, and it may also be reasonable to posit that there is no correlation between $\mathbf{x}(0)$ and $\mathbf{x}(\tau)$ other than the one introduced by constraining the starting and ending points to lie in certain macrostates. By conservation of energy, the work done by external fields is given by $W[\mathbf{x}(t)] = \beta \Delta Q[\mathbf{x}(t)] + H_{\text{sys}}(\mathbf{x}(\tau), \lambda(\tau)) - H_{\text{sys}}(\mathbf{x}(0), \lambda(0))$. Since W depends on the whole history of duration τ for the driven system and τ is assumed to be much longer than the time needed to lose all memory in
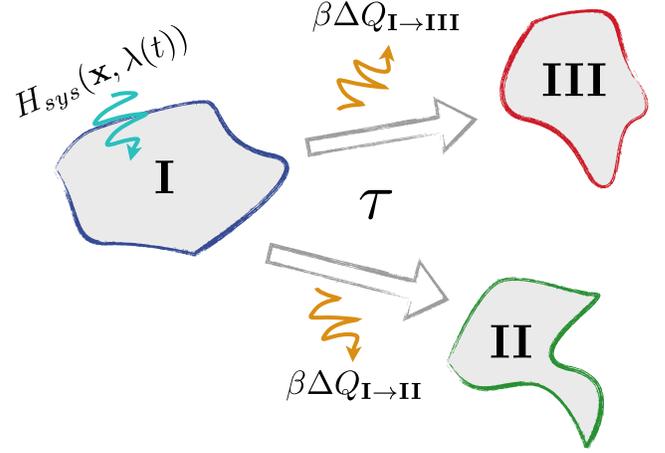

FIG. 3.   In this scenario, we consider a starting nonequilibrium macrostate **I** that is driven by an external field while it evolves stochastically in contact with a thermal bath. We may ask whether, after some finite time τ, it is more likely to be found in one of two other possible macrostates **II** and **III**. The relative likelihood of these two different possibilities will be determined in part by the statistical distributions for possible values of the external entropy production $\beta \Delta Q_{\mathbf{I} \to *}$ in each case.

the system of $p_i(\mathbf{x})$, we can treat any correlation between W and $p_f(\mathbf{x})$ or $p_i(\mathbf{x})$ as being negligible and thus write out the following separate terms:

$$\ln\left[\frac{\pi_\tau(\mathbf{I} \to \mathbf{II})}{\pi_\tau(\mathbf{I} \to \mathbf{III})}\right]$$

$$= -\ln\left[\frac{\langle \frac{p_f^{\mathbf{II}}}{p_{bz}^{\mathbf{II}}} \rangle_{\mathbf{II}}}{\langle \frac{p_f^{\mathbf{III}}}{p_{bz}^{\mathbf{III}}} \rangle_{\mathbf{III}}}\right] + \ln\left[\frac{Z_{\lambda(\tau), \mathbf{II}}^{eq}}{Z_{\lambda(\tau), \mathbf{III}}^{eq}}\right]$$

$$+ \ln\left[\frac{\pi_\tau(\mathbf{II} \to \mathbf{I})}{\pi_\tau(\mathbf{III} \to \mathbf{I})}\right] - \ln\left[\frac{\langle e^{-\beta W} \rangle_{\mathbf{I} \to \mathbf{II}}}{\langle e^{-\beta W} \rangle_{\mathbf{I} \to \mathbf{III}}}\right], \quad (8)$$

where $p_{bz}^*(\mathbf{x}(t), \lambda(t)) d\mathbf{x} = \mathbf{I}_{\mathbf{x} \in *} e^{-\beta H_{\text{sys}}(\mathbf{x}(t), \lambda(t))} d\mathbf{x} / Z_{\lambda(t), *}^{eq}$ and $Z_{\lambda(t), *}^{eq} = \int_{\mathbf{x} \in *} d\mathbf{x} e^{-\beta H_{\text{sys}}(\mathbf{x}(t), \lambda(t))}$. This result follows because, according to our assumptions and definitions, $\Delta S_{\text{tot}}^{\mathbf{I} \to *}[\mathbf{x}(t)] = \ln\{[p_i^{\mathbf{I}}(\mathbf{x}(0))]/[p_f^*(\mathbf{x}(\tau))]\} + \beta W[\mathbf{x}(t)] - \ln\{[p_{bz}^{\mathbf{I}}(\mathbf{x}(0), \lambda(0)) Z_{\lambda(0), \mathbf{I}}^{eq}]/[p_{bz}^*(\mathbf{x}(\tau), \lambda(t)) Z_{\lambda(\tau), *}^{eq}]\}$.

The first term in Eq. (8) has an appealingly intuitive meaning: The quantity $-\ln\langle p_f^*/p_{bz}^* \rangle_* \leq 0$ is a functional over the nonequilibrium distribution in macrostate $*$ that is only equal to zero when $p_f(\mathbf{x})$ is equal to the Boltzmann distribution $p_{bz}(\mathbf{x})$ [which can be shown most easily for a discrete distribution over states i using Lagrange multipliers by solving $[\partial/\partial p_f(i)] \sum_i (p_f(i)^2/p_{bz}(i) - \lambda p_f(i)) = 0$]. The behavior of this functional for small variations from p_{bz} has been examined previously in this context since one may then approximate $\ln \sum_i p_f(i)^2/p_{bz}(i) = \ln \sum_i p_f(i) e^{\ln(p_f/p_{bz})} = \ln\langle e^{\ln(p_f/p_{bz})} \rangle_f \approx \ln e^{\langle \ln(p_f/p_{bz}) \rangle_f} = \sum_i p_f(i) \ln\{[(p_f(i))/[(p_{bz}(i)]\}$, the last of which is equal to

the Kullback-Leibler divergence between the equilibrium and nonequilibrium distributions [17]. Meanwhile, $F^{eq}_{\lambda(\tau),*} = -\frac{1}{\beta}\ln Z^{eq}_{\lambda(\tau),*}$ clearly just measures the equilibrium free energy that would be obtained if the system were equilibrated at inverse temperature β while being constrained to inhabit the states belonging in macrostate $*$. Thus, we can recognize that the quite intuitive effect of these two terms together is to account for the dynamical tendency of any driven system to inhabit macrostates of low equilibrium free energy and to progress towards thermal equilibration within those macrostates. Here, we follow the example established in previous treatments of exponentially averaged work distributions [6] and, defining the so-called "dissipated work" for a trajectory ending in $*$ as $W_d = W - F^{eq}_{\lambda(\tau),*} + F^{eq}_{\lambda(0),\mathbf{I}}$, we can combine the second and fourth terms in the above equation into one path ensemble average of $\langle e^{-\beta W_d}\rangle_{\mathbf{I}\rightarrow *}$.

Thus, against the backdrop of an expected tendency to thermally equilibrate dictated by the first term, we are now in the position to weigh the effect of the latter two terms remaining in the equation. The first of these reflects the impact of kinetics. Succinctly, it says that, in a given amount of time, one is more likely to propagate to a macrostate from which one is more likely to revert back in the same amount of time. We can start to get a handle on this term by considering a simple transition-state theory model of a one-step chemical reaction between initial and final states [16]: When the activation free energy barrier is lowered, both the forward and reverse rates of the reaction are accelerated. Thus, one factor we would expect to affect the value of this term is the cumulative impact of all the kinetic barriers that need to be traversed in order to get from a given starting state to a given ending state in a fixed amount of time. Put another way, one more rapidly evolves to states that are closer to one's starting point and separated from it by fewer high barriers. Indeed, this remains the case even when we are talking about ensembles of trajectories that are likely in the forward direction; within this ensemble, the probabilities of trajectories are affected by the heights of barriers that must be traversed whether in the forward (likely) or reverse (unlikely) direction.

In light of this, we seem to be presented with two significant problems. The first of these is that we are purportedly interested in knowing how much more likely we are to see $\mathbf{II}$ happen in our driven system than $\mathbf{III}$, and yet in order to say anything concrete, it would seem that we have to know the reversal probabilities $\pi^{\mathbf{I}}_\tau$, which requires complete knowledge of all the rates of stochastic transitions that are possible in our system. Put another way, we may as well run a fully detailed microscopic simulation and just see what happens. The second problem is the well-known difficulty of averaging exponentials of large quantities: There are many microtrajectories that run from $\mathbf{I}$ to $\mathbf{II}$, and in general, it will be some subensemble of highly

improbable ones that make the dominant contribution to $\langle\exp(-\beta W_d)\rangle_{\mathbf{I}\rightarrow\mathbf{II}}$ if we average over all possible paths. This puts us in the untenable position of needing to know the dissipation associated with stochastic events in our system that are immeasurably unlikely to occur [18].

We can start at least by turning two problems into one if we note that everything we have said so far still applies if we choose not to average over *all* paths leading from one macrostate to another but rather only over a select subensemble of paths representative of most of the probability current. Our goal is to make an inference about how the probability of reaching a given macrostate is related to the *typical* thermodynamic fluxes that are associated with getting there. Thus, the practical meaning of this restriction is that one must choose a quantitative criterion that defines what we mean by "typical," which should generally be determined for the purposes of calculation by one's ability to sample the distribution of entropy production in the forward direction. Formally, we can propose a procedure whereby some finite number $\mathcal{N} \gg 1$ of paths are randomly drawn from the forward probability distribution for paths leading between the given starting and ending macrostates. Implicitly, such a procedure excludes "abnormal" types of events unlikely to occur once in $\mathcal{N}$ experiments. Thus, the associated work distribution may still be highly non-Gaussian and have substantial contributions to $\langle e^{-\beta W_d}\rangle$ from higher-order cumulants, but it will now have a value that is determined by the statistics of observable events of measurable likelihood. To present a concrete example, if we observed the same seed in the same pot of soil over the course of several weeks in a million repetitions of the same experiment, we would never expect to see a fully grown plant spontaneously self-assemble in the last ten seconds of the interval of observation; rather, all the trajectories we observed (even those that were, in some sense, "one in a million") would look like a seed growing gradually into a plant.

If we restrict the average to these "normal" forward paths, the two things that change in tandem (since we are holding the starting and ending points of the ensemble of trajectories fixed) are the dissipated work averages and the reversal probabilities: We get to average dissipation over events of measurable likelihood, and the "price" we pay is that the reversal probabilities now correspond only to the likelihoods of trajectories that look like reverse movies of trajectories that look normal in the forward direction. This method works because excluding unlikely freak occurrences from our ensemble of trajectories, by definition, does nothing to measurably alter the value of the forward probabilities (i.e., $\pi^{\text{fwd}}_{\mathbf{I}\rightarrow\mathbf{II}} \simeq \pi_{\mathbf{I}\rightarrow\mathbf{II}}$). So, for example, in the case of a seed growing into a plant, averaging $\langle e^{-\beta W_d}\rangle$ over all possible ways of going from seed to plant would give us a quantity related to the total probability $\pi_{\mathbf{II}\rightarrow\mathbf{I}}$ of turning back from plant to seed. In contrast, averaging $\langle e^{-\beta W_d}\rangle^{\text{fwd}}$ only over typical forward paths will

yield a quantity related to $\pi_{\mathbf{II}\to\mathbf{I}}^{\text{rev}} \ll \pi_{\mathbf{II}\to\mathbf{I}}$, which is the probability that we observe the plant turning back into a seed via a path that looks like a time-reversed movie of a seed growing in the normal way into a plant. We can thus denote the whole thermodynamic equation more compactly via

$$\ln\left[\frac{\pi_{\mathbf{I}\to\mathbf{II}}^{\text{fwd}}}{\pi_{\mathbf{I}\to\mathbf{III}}^{\text{fwd}}}\right] = -\Delta\ln\left\langle\frac{p_f}{p_{bz}}\right\rangle_{\mathbf{II},\mathbf{III}} + \ln\left[\frac{\pi_{\mathbf{II}\to\mathbf{I}}^{\text{rev}}}{\pi_{\mathbf{III}\to\mathbf{II}}^{\text{rev}}}\right]$$
$$- \ln\left[\frac{\langle\exp(-\beta W_d)\rangle_{\mathbf{I}\to\mathbf{II}}^{\text{fwd}}}{\langle\exp(-\beta W_d)\rangle_{\mathbf{I}\to\mathbf{III}}^{\text{fwd}}}\right]. \quad (9)$$

It should be noted in passing that in driven systems the value of the reverse probability term is not determined only by distance traveled or barriers traversed because external drives generally affect the probabilities of transition between macrostates in both forward and reverse directions. Moreover, since drives often do not impact forward and reverse transition probabilities between states in exactly the same way (since they push in particular directions at particular times) and because long paths connecting macrostates are made up of many partially independent microscopic moves, we expect $\pi_{\mathbf{I}\to*}^{\text{fwd}}$ and $\pi_{*\to\mathbf{I}}^{\text{rev}}$ between macrostates of equal equilibrium free energy to be positively, but imperfectly, correlated with each other over the range of possible choices of $*$, leading, in turn, to a range of possible values of $\langle e^{-\beta W_d}\rangle_{\mathbf{I}\to*}$ for different end states.

Taking one final heuristic step, we may also note that $-\ln\langle\exp(-\beta W_d)\rangle$ is a cumulant generating function [6] when we expand in β, and thus, it may be written as

$$-\ln\langle\exp(-\beta W_d)\rangle = \beta\langle W_d\rangle - \frac{\beta^2}{2}\sigma_{W_d}^2 + \frac{\beta^3}{6}\langle W_d\rangle_c^3\cdots$$
$$\equiv \Psi - \Phi, \quad (10)$$

where $\Psi = \beta\langle W_d\rangle$ is the average dissipated work and Φ is accordingly defined to account for all of the fluctuations about this average via $\Phi = \ln\langle\exp(-\beta W_d)\rangle + \Psi$. Because $e^{-\Psi+\Phi} = \langle\exp(-\beta W_d)\rangle \geq \exp(-\beta\langle W_d\rangle) = e^{-\Psi}$, it immediately follows that $\Phi \geq 0$, with equality only holding when W_d is δ-function distributed. Thus, Φ can be thought of as a correction due to the dispersion of the dissipated work distribution about the average Ψ that gives the heaviest weight to the leftward tail of the work distribution. This asymmetry of weighting arises because $e^{\Phi} = \langle\exp[-\beta(W_d - \langle W_d\rangle)]\rangle$ places exponentially stronger weight on lower values of W_d and thus grows even when the distribution spreads symmetrically in both directions [18].

For a simple example, one may consider the case where $p(W_d - \langle W_d\rangle)dW_d = (1/w)$ on the interval $[-w/2, w/2]$, with $w > 0$. In this instance, $\Phi = \ln[(2/\beta w)\sinh(\beta w/2)]$, which grows without bound in the limit of large $w \gg k_B T$. We can also note that the behavior of Φ is particularly

informative in the case of an ergodic nonequilibrium steady state since then $\langle\exp[-\Delta S_{\text{tot}}]\rangle = 1$, which implies that $\Psi - \Phi = 0$ for the dissipated work measured over any finite time τ. Clearly, the steady-state dissipation rate can take on any value, in principle, depending on the system in question. However, in such cases, the dispersion about the mean rate of dissipation is required by time reversibility to rise as well so that $\Psi = \Phi$. What this points to is that the appearance of Φ in our equation for relative probabilities, among other things, forces us to ignore contributions to the total average dissipated work that are associated with futile cyclical motion performed while the system dwells near a steady state. This factor is therefore closely related to the idea of "housekeeping heat" considered previously in studies of transitions between nonequilibrium steady states [8].

By incorporating this quantity into our thermodynamic expression, we may now write

$$\ln\left[\frac{\pi_{\mathbf{I}\to\mathbf{II}}^{\text{fwd}}}{\pi_{\mathbf{I}\to\mathbf{III}}^{\text{fwd}}}\right] = -\Delta\ln\left\langle\frac{p_f}{p_{bz}}\right\rangle_{\mathbf{II},\mathbf{III}} + \ln\left[\frac{\pi_{\mathbf{II}\to\mathbf{I}}^{\text{rev}}}{\pi_{\mathbf{III}\to\mathbf{I}}^{\text{rev}}}\right]$$
$$+ \Delta\Psi_{\mathbf{II},\mathbf{III}}^{\mathbf{I},\text{fwd}} - \Delta\Phi_{\mathbf{II},\mathbf{III}}^{\mathbf{I},\text{fwd}}. \quad (11)$$

Equation (11) has a rich structure and may be thought of as a generalization of the Helmholtz free energy for the stochastic evolution of arbitrarily driven nonequilibrium macrostates over finite time. Like many far-from-equilibrium thermodynamic relations, it essentially just summarizes the combined constraints on dynamics that come from time reversibility and conservation of energy and probability. What we would ultimately like to suggest here is that expressing these constraints in this form makes it possible to give them a new interpretation, in which a thermodynamic mechanism for nonequilibrium self-organization is revealed. In the next section, however, we first try to illustrate the relevance of the above expression in a few simple and instructive cases.

IV. DISSIPATIVE HISTORY AND OUTCOME-LIKELIHOOD: TWO EXAMPLES

The nonequilibrium systems described, in principle, by Eq. (11) are extremely diverse in their properties, and before discussing the general implications of the result, it will be helpful to consider how things play out in a few simple examples.

A. Driven hopping particle

First, it is appropriate to return to the hopping particle in a driven energy landscape considered in Fig. 1. Previously, we pointed out that there was a direct correlation in this system between the relative likelihood of two outcomes of the driven dynamics and the amount of heat dissipated on the way to each destination.

What we can now appreciate is that this correlation results from the weight of the $\Delta\Psi$ term in Eq. (11). Indeed, the reason this effect is so complete in this case is that our choice of external drive in fact contrived the two outcomes of x_1 and x_3 to have equal return probability: By setting the energy $E_1(t)$ to oscillate in sync with $B_{12}(t)$, the barrier height for the return jump remained the same as $B_{23} - E_3$ at all times. Moreover, in the assumed limit of transitions dominated by the lowest instantaneous energy barrier, the amount of dissipation for a jump from x_2 to x_1 is always the same ($\Phi = 0$) and always positive because $E_1 < E_2$ at the moments when $r_{2\to1}(t)$ is maximized. Referring back to Eq. (11) and noting that $\Delta\Phi = 0$ in this case, we thus can see we have recovered the predicted relationship

$$\frac{r_{2\to1}^{\max}}{r_{2\to3}} = \exp\left[\Delta\Psi\right]. \tag{12}$$

Using similar reasoning, we may also consider the effect of the fluctuations Φ on drift. The above discussion has amounted to an argument for why scenarios involving drift in a particular direction must also exhibit elevated external dissipation (that is, more positive Ψ), yet it is clearly the case that not all heat evolution in driven systems is accompanied by such statistically irreversible flow of probability density. Indeed, using the Crooks relation for microstate-to-microstate transitions in such a discrete model, one may easily show that when $\exp[-\beta\Delta Q]$ is averaged over all stochastic trajectories that both start and end in the same state x_i, it is generally the case that $\Psi_{i\to i} = \Phi_{i\to i} \geq 0$, reminding us that much of the dissipation generated in a driven system can result from futile cycles that go around in circles rather than drifting in some new direction in phase space [8]. Thus, some ways of making Ψ more positive also bring about an increase of Φ in lockstep that cancels any contribution to the flow of probability in Eq. (11). Put another way, high fluctuations in entropy production lessen the degree to which drift and dissipation are correlated.

To see this effect of fluctuations in the case of our hopping particle, we can specialize to a slightly different scenario (Fig. 4). Now considering a pair of states x_1 and x_2, we can initialize the particle in state x_2 with constant energy $E_2 = 0$ and suppose there are two distinct paths that connect the states. The first path has a barrier height $B_{12}(t) = \Delta E + \Delta E\cos(\omega t)/2$, whereas the second path has a barrier of height $B_{12}^*(t) = \Delta E - \Delta E\cos(\omega t)/2$, and both barriers are governed by the same time constant r. The energy of state x_1, meanwhile, is assumed to vary with time as $E_1(t) = \Delta E\cos(\omega t)/2$.

In a system thus described, we can straightforwardly compute all of the relevant quantities if we assume we are in a regime where $\beta\Delta E \gg 1$ and $r \ll 1/\tau \ll \omega$ so that transitions are rare events that take place during the moments in the drive cycle of their maximum
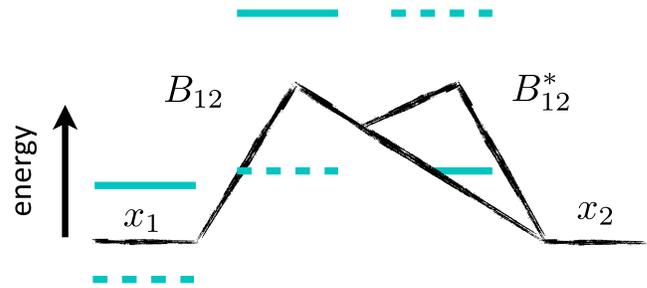

FIG. 4.   In the driven scenario depicted schematically here, two states x_1 and x_2 are separated by two different barriers representing two different possible paths for transiting between the states. Because the barriers are driven out of phase with each other, and since $\beta\Delta E \gg 1$ and $r \ll 1/\tau \ll \omega$, the return probability and average dissipation for transitions from x_2 to x_1 are independent of ΔE, the choice of which should still affect the drift rate from x_2 to x_1 as well as the fluctuations Φ for the pair of paths. As in Fig. 1, the energies of the driven states are drawn at their extremal values in blue, with dashed lines and solid lines, respectively, in phase with each other.

likelihood. In this case, it is easy to see that the particle is equally likely to first jump to x_1 via either barrier since $\max(r_{2\to1}(t)) = \max(r_{2\to1}^*(t)) = r\exp[-\beta\Delta E/2]$. Interestingly, however, the mean dissipated work averaged over the weights of these paths should be $\Psi = \beta(\Delta E/2 - \Delta E/2)/2 = 0$, due to the fact that events on one path happen in phase with the rise of E_1 and events on the other path happen π radians out of phase with it. Moreover, in the limit of ΔE we have chosen, the total return probability $r\tau = \mathrm{const}$ is effectively insensitive to the exact value of the barrier height since the return journey will always be made via B_{21}^* and not B_{21}. However, in this same regime, changes in ΔE *strongly* affect both the forward transition rate and the fluctuations in entropy production; specifically, the transition rate is $\ln\overline{r_{2\to1}^{\mathrm{tot}}} = \ln[2re^{-\beta\Delta E}I_0(\beta\Delta E/2)] \simeq -\beta\Delta E/2$, and the fluctuations are $\Phi = 0 - (-\ln\langle e^{-\Delta S}\rangle) = \ln\cosh[\beta\Delta E/2] \simeq \beta\Delta E/2$. Thus, for this particular case, we may write what we should have expected from Eq. (11):

$$\ln\overline{r_{2\to1}^{\mathrm{tot}}} \simeq -\Phi, \tag{13}$$

from which it is apparent that, as fluctuations in entropy production rise with return probability and average dissipation already held fixed, the forward probability rate must fall sharply. The origin of this effect lies in the diversity of paths connecting states, as well as the resulting possibility that the drive can sometimes *help* one return to the starting point via a different path than it pushed one along when the journey first began.

B. Competing self-replicators

In order to get a more phenomenological perspective on how dissipative history can be predictive of outcomes in

nonequilibrium dynamics, we can now consider two populations of self-replicating particles, A and B. The coarse-grained states of interest correspond to different ordered pairs (N_A, N_B) which count the numbers of individuals of each type in the system. In any infinitesimal interval of duration dt, an individual of type A or B can spontaneously and instantaneously divide into two individuals via a highly irreversible process $X \rightarrow X + X$ with a probability of $g_A dt$ or $g_B dt$, respectively. Without loss of generality, we will set $g_B > g_A$ so that one of these exponentially growing populations is expected, on average, to become an ever-increasing fraction of the total population.

We further assume that both types of self-replicators are equally likely to undergo a reversal reaction $X + X \rightarrow X$ with the same stochastic rate constant δ. Although the probability for this process to happen in reverse is assumed to be very small, the time reversibility of the underlying physics forbids it from being zero. Under these rules, the probability of finding N individuals of type X in a given population at time t obeys the master equation

$$\dot{p}_{N_X}(t) = -\left(gN_X + \delta\frac{N_X(N_X - 1)}{2}\right)p_{N_X}(t)$$
$$+ g(N_X - 1)p_{N_X-1}(t)$$
$$+ \delta\frac{(N_X + 1)N_X}{2}p_{N_X+1}(t). \qquad (14)$$

We consider the limit where $\delta \rightarrow 0$, but δ must remain nonzero in order for the entropy production

$$\Delta S_X = -\ln\frac{p[N_X \rightarrow N_X - 1]}{p[N_X - 1 \rightarrow N_X]} = \ln\left[\frac{2g_X}{\delta}\right] - \ln N_X \qquad (15)$$

associated with each replication event to remain finite [14] (Note: The form of the entropy production expression takes a slightly different form here than in Ref. [14] because the reverse path defined here occurs by pair annihilation, rather than single-particle decay). We therefore restrict our attention to the case where δ is nonzero but exponentially small compared with the growth rate: $\ln(g_X/[\delta N_X]) \gg 1$.

Our goal here is to illustrate that external entropy production becomes predictive of the stochastic dynamics of this system once comparisons are made between possible outcome states that all have the same value (or greater) of spontaneous reversal probability. It must be emphasized that at no point is it assumed here that this reversal probability is constant across all possible choices of ordered pairs (N_A, N_B). On the contrary, this quantity is expected to vary dramatically depending on how N_A and N_B are chosen. As we shall see, however, the examination of a large set of such ordered pairs that share the same value of this reversal term will allow us to make a correct inference about an emergent quality in the system's

dynamical behavior as a whole from the way entropy production depends on N_A and N_B.

Since δ is negligible for the purposes of solving Eq. (14) in the regime we are considering, the probability for stochastic evolution forward in time for both species from a starting condition of $(N_A, N_B) = (1, 1)$ over time τ is given by

$$p[(1, 1) \rightarrow (N_A, N_B), \tau] = e^{-g_A\tau}(1 - e^{-g_A\tau})^{N_A-1}$$
$$\times e^{-g_B\tau}(1 - e^{-g_B\tau})^{N_B-1}, \qquad (16)$$

which may be shown by inspection to satisfy the master equation governing the whole population in the $\delta \rightarrow 0$ limit. Our goal, however, is to compute the reversal probabilities for returning from a given final state with population numbers (N_A, N_B) to an initial state with one individual of each species, in a fixed time interval of equal duration τ. The most straightforward way of doing this in this particular system is to leave all the different contributions to entropy together in one bundle calculated from Eq. (15) and exploit the Crooks-type relation in Eq. (7) that relates entropy production to forward and reverse probabilities. In this model, the total entropy production caused by a population as it goes from one individual of each species to N_A and N_B of them, respectively, will be assumed as

$$\ln\frac{p[(1, 1) \rightarrow (N_A, N_B), \tau]}{p[(N_A, N_B) \rightarrow (1, 1), \tau]}$$
$$= \Delta S[(1, 1) \rightarrow (N_A, N_B)]$$
$$= \sum_{X=\{A,B\}}\left[(N_X - 1)\ln\left[\frac{2g_X}{\delta}\right] - \ln N_X!\right], \qquad (17)$$

which accounts both for the internal entropy of the mixture of indistinguishable self-replicators and for the heat dissipation in the surrounding bath. Combining this expression with the one for $p[(1, 1) \rightarrow (N_A, N_B), \tau]$ enables us to write the reversal probability in time τ as

$$p[(N_A, N_B) \rightarrow (1, 1), \tau]$$
$$= e^{-g_A\tau}(1 - e^{-g_A\tau})^{N_A-1}$$
$$\times e^{-g_B\tau}(1 - e^{-g_B\tau})^{N_B-1}\left(\frac{\delta}{2g_A}\right)^{N_A-1}$$
$$\times \left(\frac{\delta}{2g_B}\right)^{N_B-1}N_A!N_B!. \qquad (18)$$

Although we have obtained this result by assuming thermodynamic consistency, one may confirm that it is correct by computing the quantity directly as a path integral over individual microtrajectories.

All ordered pairs $(N_A \geq 0, N_B \geq 0)$ are, in principle, possible outcomes of this process after a time τ, and under
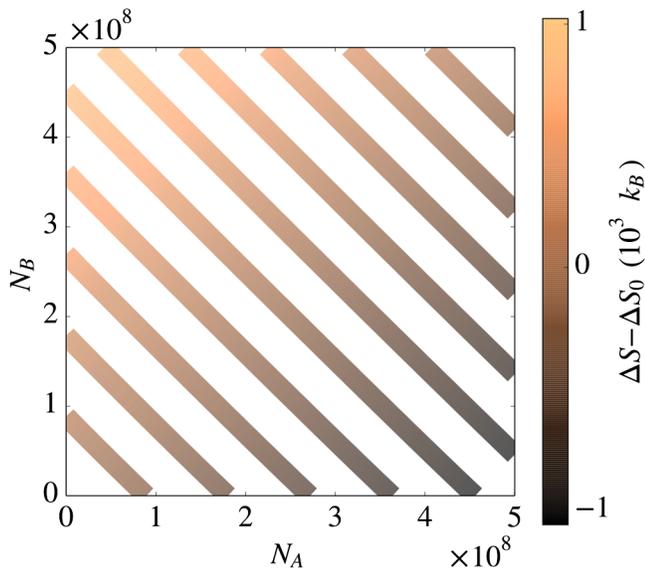
FIG. 5. Lines of fixed reversal probability based on Eq. (18), with $g_A = 1$, $g_B = 2$, $\tau = 10$, and $\delta = e^{-100}$. Each line is colored to show the entropy produced while generating each ending state from the initial state $N_A = N_B = 1$, relative to that of the state on the line where $N_B = N_A$. On each of these surfaces of fixed reversal probability, the most likely outcome by orders of magnitude is one where $N_B \gg N_A$.

the limits we have described, their associated return probabilities vary by many orders of magnitude (Fig. 5). There are, however, lines through this state space on which the reversal probabilities are constant, and we can use Eq. (18) to obtain the relation between N_A and N_B that defines these lines. We are interested in this relation in the limit of vanishingly small δ ($\ln(g_X/[\delta N_X]) \gg 1$), $N_X \gg 1$ and comparable order-of-magnitude sizes for $g_A \sim g_B \sim (1/\tau) \ln N_X$. In this case, by taking the logarithm of Eq. (18) and setting it equal to a constant, C, we find that the dominant contribution comes from the $N_X \ln \delta$ terms, and we can write $N = N_A + N_B \simeq C/\ln \delta$, so fixed-reversal probability lines nearly become lines of fixed total population number. In fact, this correspondence cannot be exact since the discreteness of integer lattice sites prevents most states from having identical values of reversal probability. We can, however, use the collection of lattice sites with fixed particle number N as a proxy for the fixed return probability ensemble in its role as a large and diverse set of outcomes. On such level sets, the entropy production takes the simple form

$$\Delta S = C' + \ln p[(1,1) \rightarrow (N_A, N_B), \tau]$$
$$\approx C'' + N_B(e^{-g_A \tau} - e^{-g_B \tau}), \qquad (19)$$

where C', C'' are constants independent of N_A and N_B.

Figure 5 shows how ΔS spans a wide range of values over each line of fixed reversal probability for a typical choice of parameters. Since $g_B > g_A$, the expression in parentheses is always positive, so entropy production increases with increasing N_B along each line. It further follows that the relative likelihood of states on this level curve increases exponentially with N_B, so we obtain here the result for this model that in any half-plane bounded from above by some value of reversal probability, the most likely outcomes (which are also those that have produced the most entropy) are those where $N_B \gg N_A$.

It is worth restating the flow of the above argument. Of course, we already find it obvious that in a competition between two exponentially growing populations of individuals that essentially never die, the type of individual that grows at a faster rate is expected to become more numerous with the passage of time. This intuition is all we need to understand the dynamics of this particular system. However, our goal here is to provide an explanation for this outcome in a thermodynamic language that will remain applicable in settings where the stochastic dynamics are much less straightforward to predict by other means. Here, total entropy production was determined simply by the number of self-replication events of each type, whereas reversal probability essentially tracks the total population size. Once we bound this latter quantity from above, we find that the likeliest outcomes (which are those for which $N_B \gg N_A$) always turn out to be those for which ΔS is most positive, *given* the reversal constraint.

Conversely, we can say that if we had no *a priori* intuition for how the system should behave, we could still infer the likely takeover by species B from the fact that ΔS increases with N_B along lines of fixed reversal probability. Here the "structural quality" that correlates strongly with an exceptional dissipative history on any fixed reversal probability surface is having $N_B \gg N_A$, and it is this quality that indeed is expressed in the outcome of the stochastic dynamics. Moreover, if we instead looked at surfaces of fixed entropy production, we would uncover an inverse relationship between reversal probability and total particle number. Accordingly, although an infinite number of different particle numbers are possible, in principle, at any finite time, we expect (and find) that sufficiently large values of N are less likely to be observed than smaller ones.

The examples we considered here effectively illustrate why time reversibility in thermally fluctuating systems implies a general relationship between the concerted flow of probability density and the consistent conversion of work to positive amounts of dissipated heat. In the remainder of this article, we discuss how this relationship may be construed to suggest a thermodynamic mechanism for adaptive self-organization.

V. DISCUSSION

In this article, we have demonstrated a general expression for the time evolution of the probability distribution over macrostates for a driven classical many-body system. Moreover, we interpreted this equation to be a summary of

the distinct pressures pushing the system towards particular outcomes. The first such pressure is the most obvious one: All things being equal, we expect our collection of matter in contact with a heat bath to tend towards thermal equilibrium. Next, we account for kinetic accessibility: Not all arrangements of a system are equally "adjacent" or "close" to a given starting point, and thus, only certain arrangements should be considered equally accessible on any finite time scale of evolution. Beyond these effects, the rest of what determines relative likelihood is a function of a state's implicit history of thermodynamic flux. Outcomes from absorption and dissipation of work that is consistently high (corresponding to exceptionally positive Ψ and small Φ) are expected to be favored once the kinetic and equilibration terms are held fixed in an all-things-equal comparison.

It was easiest to see how such a general thermodynamic statement played out in the two simple examples we already considered, but now we may ask whether the same theoretical framework might reveal interesting consequences in more complex nonequilibrium settings. To answer this question, we pose a thought experiment that will allow us to see implications for systems that seem to undergo qualitative transformations in their pattern of organization as a result of strong external driving.

Imagine an experimental realization of a process of stochastic evolution for a many-particle system in which some particular outcome called **II** is reliably the one observed after the system is initialized in some nonoverlapping state **I** and driven for finite time τ with some particular pattern of time variation in the external field $\lambda(t)$. For a concrete example, one might consider the phenomenon of shear thickening in a driven colloid, where particle clumping is caused by the application of external shear to a formerly dispersed mixture [19,20]. By requiring that $\lambda(0) = \lambda(\tau)$ and that $F_{\mathbf{I}}^{eq} \leq F_{\mathbf{II}}^{eq}$, we avoid the trivial cases where the transition in question would be favored by undriven dynamics of equilibration. Moreover, we assume that, while the forward probability of evolution $\pi_{\mathbf{I} \to \mathbf{II}}^{\text{fwd}}$ is close to 1, the reverse probability $\pi_{\mathbf{II} \to \mathbf{I}} \ll 1$ is very small (which, in turn, usefully implies that $\pi_{\mathbf{II} \to \mathbf{I}}^{\text{rev}} < \pi_{\mathbf{II} \to \mathbf{I}} \ll 1$ is also very small), so the effect of the drive is not simply to symmetrically accelerate the kinetics of exchange between **I** and **II**. Lastly, we need to posit that, with the driving field turned off, the forward transition probability $\pi_{\mathbf{I} \to \mathbf{II}}^{\text{fwd},eq}$ should be much less than what results in the presence of the drive. Taken together, these stipulations mean that we are aiming specifically to consider a system where introducing the time-varying external driving field triggers a reliable and irreversible configurational shift into a state that would not normally be explored as a result of thermal fluctuations alone in the same amount of time.

In the scenario thus described, the assumptions we have already stated are sufficient to imply that **II** will have the optimally positive value of $\Psi - \Phi$ across a large and diverse set of possible macroscopic arrangements for the system.

Of course, not all arrangements of a many-body system will be equally kinetically trapped; thus, not all macrostates we construct will satisfy our assumption of $\pi_{\mathbf{II} \to \mathbf{I}}^{\text{rev}} < \pi_{\mathbf{II} \to \mathbf{I}} \ll 1$. However, in a system containing many different interacting degrees of freedom, the space of possible particle arrangements that do satisfy this inequality should be expected to remain combinatorially vast, and it is over these possible outcomes that the overwhelmingly likeliest one (**II**) must have an exceptional implicit history of consistently positive dissipative flux. (Note: Both the focus on dissipative *history* and account of the effect of fluctuations through the contribution of the Φ term serve to distinguish the result discussed here from maximum entropy production principles that have been explored elsewhere [21].)

What does having an exceptionally dissipative past mean for the way the system is configured in the present? To provide an answer, it is essential to realize that not every microscopic arrangement **x** available to our system of interest will absorb the same amount of work from the same time-varying field; indeed, the instantaneous work rate of the field at any given moment is $\dot{W} = \partial H_{\text{sys}}(\mathbf{x}, \lambda(t))/\partial t$, which generally must depend on the value of **x** [6,7]. This point can be made most intuitive by considering a system with a rugged energy landscape characterized by the existence of many degenerate local minima that might act as kinetic traps below a certain temperature. Small deviations from one such local minimum should behave harmonically and give rise to a spectrum of corresponding normal-mode frequencies specific to the local steepness of the potential. Consequently, when assessing how much work will be absorbed by the system when it is, for example, sinusoidally driven by an external force with frequency ω, a pertinent question will be the following: How many normal modes for the system have natural frequencies close to ω, such that they might exhibit a high-amplitude, high-dissipation response to the drive? This number of such modes could vary substantially across the vast space of many-particle arrangements. Thus, from such an example, we can take the more generic intuition that special subsets of particle configurations may be particularly well (or poorly) suited to absorb work from a given drive.

We expect, then, that in many cases where external driving brings about a strong shift in a system's state of organization, the outcome we observe at long times should reflect that the system had to be specially configured at some time in its past in order to absorb and dissipate large amounts of work from the drive. Of course, exactly how such a tendency toward "dissipative adaptation" to the drive plays out could lead in very different directions depending on the details of the system. For example, in a scenario where absorption and dissipation of work from a drive fuels a process of positive feedback (such as in self-replication), one expects the system to exhibit an improving ability to cause dissipation in its surroundings over time. On the other hand, in scenarios where the absorption and dissipation of

work causes highly irreversible structural changes that cause a drop in the rate of work absorption (such as when resonant vibrations with a song cause the shattering of a wine glass into shards that no longer resonate well), one could observe an end state with a lessened ability to absorb work. In either event, however, we expect that the eventual configuration of the system should be correlated in an informative way with the moments in its history when it absorbed large amounts of work in order to undergo highly irreversible configurational changes.

Strikingly, this manner of thinking can be thought of as the natural generalization of the way people often reason about evolutionary adaptations in Darwinian examples in biology: When a self-replicator exhibits adaptive traits that match well with its present environment, we can seek an explanation in the hypothesis that the organism's ancestors possessed these traits and found them helpful in the struggle to survive and generate many offspring [22]. As we have already seen in the previous section, the link between this picture and the thermodynamic one comes from the fact that, as a result of its statistical irreversibility, self-replication is necessarily a dissipative process [14], so exponential population growth turns out to be one very reliable mechanism of driving external entropy production. In this more general physical frame, however, we can appreciate that even a system devoid of any recognizable self-replicator might still end up seeming "well adapted" to environmental drives in light of the marks on its current configuration left by its exceptional dissipative history.

VI. CONCLUSION

It has certainly long been known that an endless variety of far-from-equilibrium many-particle Newtonian systems are capable of exhibiting self-organization phenomena in which strikingly patterned structures emerge in the presence of dissipative external drives [5]. Whether in sand dunes or snowflakes, in hurricanes or in spiral bundles of protein filaments and motors [3,4], the nonequilibrium world offers many test cases for the general hypothesis that organized, kinetically stable structures emerge and persist because their formation is accompanied by extra work absorption and dissipation [1,2]. While the story in each of these cases must be different in many of its details, we may speculate that a thermodynamic commonality would be revealed to underlie all of them if the right physical observables were tracked and compared.

Guided by these theoretical considerations, we recently demonstrated the predicted phenomenon in a simple simulation framework that tracks the vibration spectrum of a sinusoidally driven toy chemical mixture over time. The main result of the study (which will be published in a separate article, rather than shown here) is that we do indeed see emergent "adaptive" resonance in the system, such that our choice of frequency for an external driving field determines the location of the peak in the resonance

spectrum for a particle mixture that evolves stochastically in the presence of such an environmental drive. This finding turns out to be highly suggestive of the results of recent experiments performed on silver nanorods assembling in the presence of light fields of different colors [23], which similarly "learned" to match their surface plasmon resonance to the frequency of the driving field. Thus, we are encouraged to explore further with related models, spurred on by the intriguing possibility that lifelike adaptive behavior in nonequilibrium systems may turn out to be surprisingly common, now that we have begun learning how to look for its physical signatures.

ACKNOWLEDGMENTS

———————

[1] M. C. Cross and P. C. Hohenberg, *Pattern Formation outside of Equilibrium*, Rev. Mod. Phys. **65**, 851 (1993).

[2] J. P. Gollub and J. S. Langer, *Pattern Formation in Nonequilibrium Physics*, Rev. Mod. Phys. **71**, S396 (1999).

[3] V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, *Polar Patterns of Driven Filaments*, Nature (London) **467**, 73 (2010).

[4] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, *Spontaneous Motion in Hierarchically Assembled Active Matter*, Nature (London) **491**, 431 (2012).

[5] I. Prigogine and G. Nicolis, *Biological Order, Structure and Instabilities*, Q. Rev. Biophys. **4**, 107 (1971).

[6] C. Jarzynski, *Nonequilibrium Equality for Free Energy Differences*, Phys. Rev. Lett. **78**, 2690 (1997).

[7] G. E. Crooks, *Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences*, Phys. Rev. E **60**, 2721 (1999).

[8] T. Hatano and S.-i. Sasa, *Steady-State Thermodynamics of Langevin Systems*, Phys. Rev. Lett. **86**, 3463 (2001).

[9] U. Seifert, *Stochastic Thermodynamics: Principles and Perspectives*, Eur. Phys. J. B **64**, 423 (2008).

[10] M. Esposito and C. Van den Broeck, *Three Detailed Fluctuation Theorems*, Phys. Rev. Lett. **104**, 090601 (2010).

[11] C. Jarzynski, *Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale*, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[12] G. Verley, R. Chétrite, and D. Lacoste, *Inequalities Generalizing the Second Law of Thermodynamics for Transitions between Nonstationary States*, Phys. Rev. Lett. **108**, 120601 (2012).

[13] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, *Thermodynamics of Prediction*, Phys. Rev. Lett. **109**, 120604 (2012).

[14] J. L. England, *Statistical Physics of Self-Replication*, J. Chem. Phys. **139**, 121923 (2013).

[15] A. C. Barato and U. Seifert, *Thermodynamic Uncertainty Relation for Biomolecular Processes*, Phys. Rev. Lett. **114**, 158101 (2015).

[16] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, New York, 2003).

[17] D. A. Sivak and G. E. Crooks, *Near-Equilibrium Measurements of Nonequilibrium Free Energy*, Phys. Rev. Lett. **108**, 150601 (2012).

[18] C. Jarzynski, *Rare Events and the Convergence of Exponentially Averaged Work Values*, Phys. Rev. E **73**, 046105 (2006).

[19] D. R. Foss and J. F. Brady, *Structure, Diffusion and Rheology of Brownian Suspensions by Stokesian Dynamics Simulation*, J. Fluid Mech. **407**, 167 (2000).

[20] X. Cheng, J. H. McCoy, J. N. Israelachvili, and I. Cohen, *Imaging the Microscopic Structure of Shear Thinning and Thickening Colloidal Suspensions*, Science **333**, 1276 (2011).

[21] L. M. Martyushev, *The Maximum Entropy Production Principle: Two Basic Questions*, Phil. Trans. R. Soc. B **365**, 1333 (2010).

[22] M. A. Nowak, *Evolutionary Dynamics* (Harvard University Press, Cambridge, MA, 2006).

[23] S. Ito, H. Yamauchi, M. Tamura, S. Hidaka, H. Hattori, T. Hamada, K. Nishida, S. Tokonami, T. Itoh, H. Miyasaka, and T. Iida, *Selective Optical Assembly of Highly Uniform Nanoparticles by Doughnut-Shaped Beams*, Sci. Rep. **3**, 3047 (2013).