

Fluctuating Landscapes and Heavy Tails in Animal Behavior

Antonio Carlos Costa ^{1,*}, Gautam Sridhar ², Claire Wyart ² and Massimo Vergassola ¹

¹Laboratoire de Physique de l'École normale supérieure (ENS), Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France

²Sorbonne University, Paris Brain Institute (ICM), Inserm U1127, CNRS UMR 7225, Paris, France



(Received 23 January 2023; accepted 6 March 2024; published 2 April 2024; corrected 15 April 2024)

Animal behavior is shaped by a myriad of mechanisms acting on a wide range of scales, which hampers quantitative reasoning and the identification of general principles. Here, we combine data analysis and theory to investigate the relationship between behavioral plasticity and heavy-tailed statistics often observed in animal behavior. Specifically, we first leverage high-resolution recordings of *Caenorhabditis elegans* locomotion to show that stochastic transitions among long-lived behaviors exhibit heavy-tailed first-passage-time distributions and correlation functions. Such heavy tails can be explained by slow adaptation of behavior over time. This particular result motivates our second step of introducing a general model where we separate fast dynamics on a quasistationary multiwell potential, from nonergodic, slowly varying modes. We then show that heavy tails generically emerge in such a model, and we provide a theoretical derivation of the resulting functional form, which can become a power law with exponents that depend on the strength of the fluctuations. Finally, we provide direct support for the generality of our findings by testing them in a *C. elegans* mutant where adaptation is suppressed and heavy tails thus disappear, and recordings of larval zebrafish swimming behavior where heavy tails are again prevalent.

DOI: [10.1103/PRXLife.2.023001](https://doi.org/10.1103/PRXLife.2.023001)

I. INTRODUCTION

Animals continuously sense, process sensory information, and respond appropriately to ensure survival. High-dimensionality and multiple timescales of these far-from-equilibrium systems challenge quantitative understanding. Yet, recent advances in machine vision technologies (e.g., [1–3]) make it possible to record an animal's pose in unconstrained environments with unprecedented resolution. Such data now span several orders of magnitude [4], motivating modeling approaches that can bridge from subsecond movements to hours-long strategies.

Despite these technical advances, a complete microscopic description is not available and, most likely, out of reach. Indeed, that would require the current posture of the animal together with its physiological, sensory, and motor state: the uncountable number of molecules involved makes it unrealistic to track them all. Progress relies on the educated hope that so many details are not needed, as selected statistical physics examples illustrate [5]. To wit, an effective equation for the slowly varying concentration field is sufficient to capture how odor molecules diffuse in the air [6]. Much of statistical mechanics relies on the identification of

such slowly varying macroscopic modes, which, through a timescale separation, depend only statistically on microscopic details. Identifying macroscopic modes may not be a simple task, though, and it requires intuition often not immediate for far-from-equilibrium systems as encountered in biology. Here, we leverage the notion of slowly varying collective variables to motivate our introduction of reduced-order models directly from imaging data of behaving animals.

Our starting point is the nematode *Caenorhabditis elegans*, a pivotal model organism [7,8]. On a two-dimensional agar plate, worms move by propagating dorsoventral waves throughout their bodies and controlling their frequency, wavelength, and direction to move forward, backward, or turn. Long sequences of such short-lived movements exhibit signatures of chaos [9,10]. Despite this inherent variability, time-delay embedding [11–15] yields a high-fidelity Markov model that predicts *C. elegans* foraging behavior [16]. The resulting simulated worms are nearly indistinguishable from real ones across a wide range of scales. This Markov model also directly recovered long-lived metastable states that correspond to transitions between relatively straight paths (“runs”) and not-so-abrupt reorientations (“pirouettes”) [16,17] (akin to the run-and-tumbling of bacteria [18]), thus providing an effective coarse-grained description of the dynamics.

Empirical evidence for the emergence of stereotypy in the dynamics of *C. elegans* reflects the timescale separation between short-term movements in a given behavioral state and long-term transitions between states. Here, we make this evocative picture concrete by recasting it in terms of models of particles hopping among wells in a potential landscape. In the first section, we build an effective Langevin description for the inferred “run-and-pirouette” dynamics. Notably, we find

*Present address: Sorbonne University, Paris Brain Institute (ICM), Inserm U1127, CNRS UMR 7225, Paris, France.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

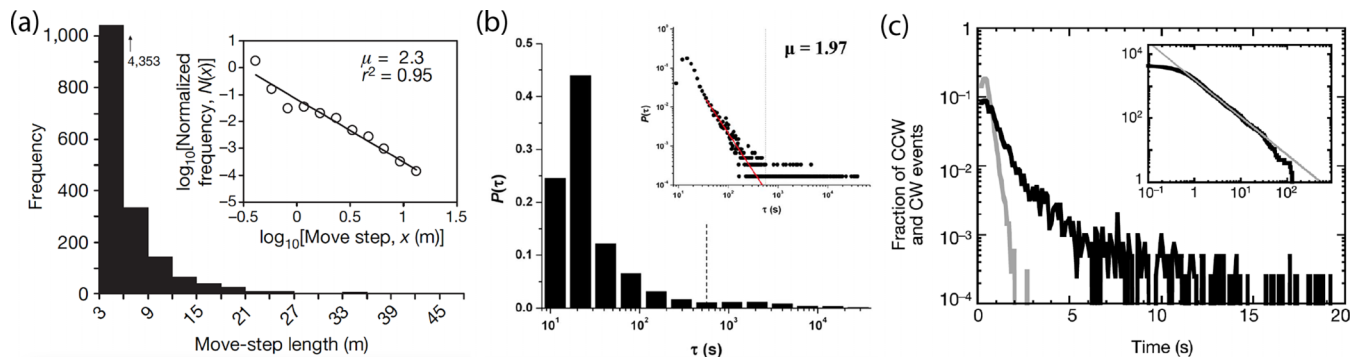


FIG. 1. Power-law distributions observed across species. (a) Distribution of step lengths ℓ for an individual basking shark (*Cetorhinus maximus*) (adapted from Ref. [19]). The inset shows the probability density $f(\ell)$ in a log-log scale, and the power-law fit $\ell^{-\mu}$ with $\mu = 2.3$. Assuming a constant speed, the time spent in a step would also be distributed as $f(t) \approx t^{-\mu}$. (b) Distribution of the time between decisions in a choice task by Sprague Dawley rats (adapted from Ref. [20]). The inset shows the same curve in log-log scale, and the power-law fit $t^{-\mu}$ with $\mu = 1.97$. (c) Distribution of the duration of clockwise (gray) and counterclockwise (black) rotations of a single *Escherichia coli* motor (adapted from Ref. [21]). The inset shows the complementary cumulative distribution function (black) with a superposed power law t^{-1} (gray), corresponding to a probability density $\sim t^{-2}$.

long-range correlations and heavy-tailed distributions of residence times spent in the two states (instead of the exponential expected for independent transition events). Our observation adds to the body of evidence showing that times spent in a given behavioral state are often heavy tailed. Indeed, power-law distributions $f(t) = t^{-2}$ [where $f(t)$ is the probability density of observing a residence time of duration t] are found extensively across species (see Fig. 1). In the context of search behavior, such observations have led to the hypothesis that Lévy flights (with an exponent -2) result in efficient search strategies and are then evolutionarily selected [22–26].

Our goal here is to combine data analysis and theory to show that heavy-tailed distributions can emerge from a slow adaptation of behavior over time. First, we infer time-dependent model parameters from the worm data and show that the observed heavy-tailed distributions can be explained by slow adaptation. Then, we introduce a general model and analyze it theoretically to account for the ubiquitous observation of heavy-tailed statistics in animal behavior. The model features potential landscapes that slowly fluctuate in time, and we demonstrate heavy-tailed first-passage times and long-range correlations. The specific point that we bring here is that we obtain an analytical expression for the exponents of the power-law distributions as a function of the strength of the fluctuations. The scaling t^{-2} mentioned above is recovered as a special limiting case. Finally, the generality of our point that behavioral plasticity may be responsible for heavy tails is strengthened by the analysis of experiments on a *C. elegans* mutant and on larval zebrafish that confirm our predictions.

II. DATA-DRIVEN ANALYSIS REVEALS HEAVY TAILS IN *C. ELEGANS* BEHAVIOR: THE ROLE OF ADAPTATION

We leverage a previously analyzed dataset in which 12 laboratory-strain N2 worms are placed on an agar plate and allowed to freely explore for $T_{\text{expt}} = 35$ min. [27]. Our procedure is illustrated in Fig. 2 and summarized hereafter, with more details deferred to Appendix A.

From each video frame (sampled every $\delta t = 1/16s$), we extract the worm’s centerline, measure tangent angles equally spaced along the body, and subtract the overall rotation of the worm to obtain the animal’s posture vector θ_t . Given a time t , the future evolution of θ does not depend on θ_t only, which reflects the effect of history and breaks Markovianity. This problem was circumvented in Refs. [9,16] by including past postures in the description of the system. In other words, short-term memory is taken into account by expanding the state space so that it admits an approximately Markovian description. The procedure detailed in Refs. [9,16] yields that a sequence $X_{K^*}(t) = \{\theta_{t-K^*\delta t}, \dots, \theta_{t-\delta t}, \theta_t\}$ of $K^* = 11$ stacked postures is sufficient to determine future statistics. The corresponding probability density $\rho_t = \rho(X_{K^*}, t)$ is advanced in time by the so-called transfer operator \mathcal{L} :

$$\frac{d}{dt}\rho_t = \mathcal{L}\rho_t, \quad (1)$$

which does not need more specifics here.

While the θ variables are continuous, it is more efficient to cluster the space of posture stacks X_{K^*} . Clustering yields a set of discrete states that summarize information on the dynamics. The operator \mathcal{L} in Eq. (1) reduces then to a matrix with diagonal and/or off-diagonal entries expressing the probability to remain in the current discrete state or jump to another one. The resulting Markov chain is embodied in the transition matrix $P_{ij}(\tau) = (e^{\mathcal{L}\tau})_{ij}$, which expresses the probability of transitioning between discrete states s_i to s_j in a time τ . The procedure is detailed in Refs. [10,16] (see also Appendix A).

To conclude the description of our data analysis, we are left to notice that eigenvalues Λ_i and eigenvectors ϕ_i of the matrix \mathcal{L} and its exponential $e^{\mathcal{L}\tau}\phi_i = e^{\Lambda_i\tau}\phi_i$ provide a hierarchy of dynamical timescales [29–31]. For a mixing system, there is a unique largest eigenvalue $\Lambda_1 = 0$ that corresponds to the steady-state ϕ_1 . The remaining eigenfunctions $\phi_{i>1}$, ranked by their decreasing real parts, correspond to collective variables that relax on faster and faster timescales $\sim |\text{Re}(\Lambda_{i>1})|^{-1}$.

For the measured *C. elegans* foraging dynamics, the eigen-spectrum of \mathcal{L} reveals a slow mode ϕ_2 that is relatively well separated from the rest of the eigenmodes and coarse-grains

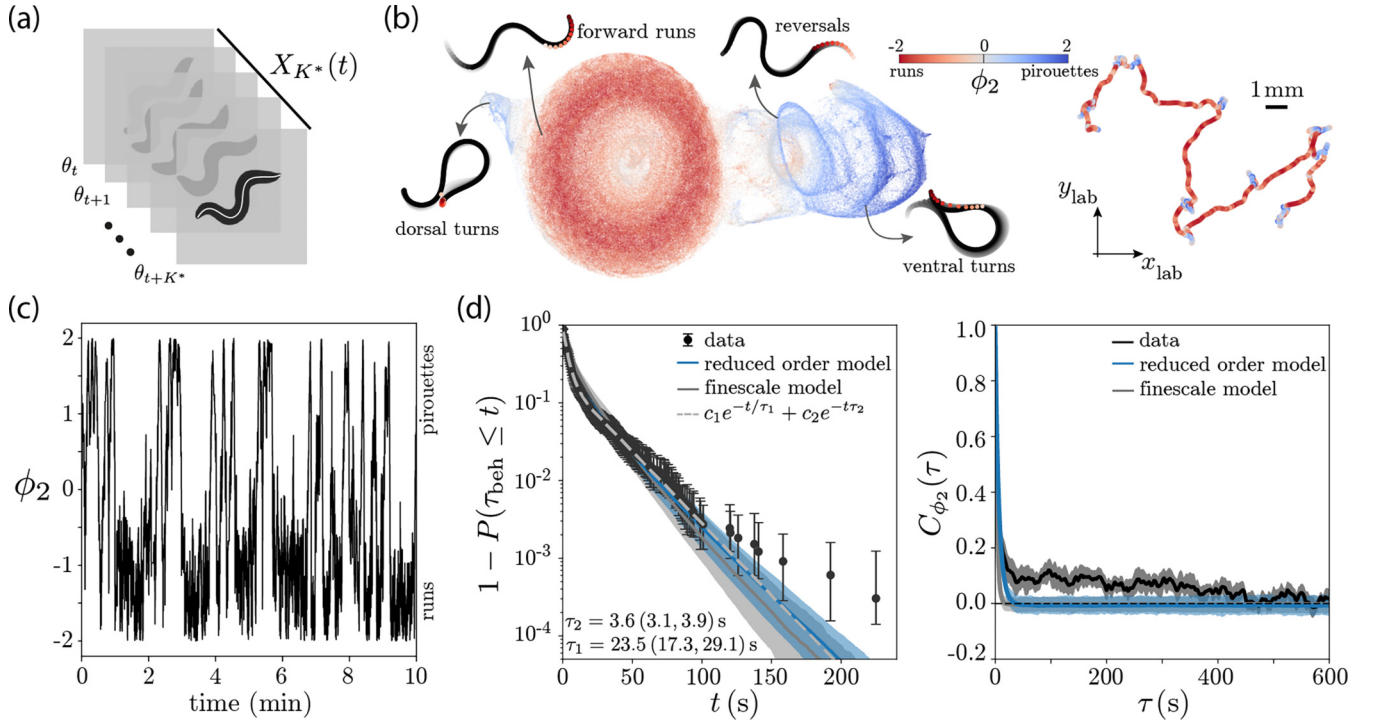


FIG. 2. A reduced-order model of *C. elegans* foraging dynamics. (a) From video imaging data, we measure local tangent angles along the body to obtain the body posture vector θ_t , as in Ref. [28]. A series of K^* such vectors are then stacked to yield the variable $X_{K^*}(t)$ defined in the text, which captures short-term memory. (b) A high-fidelity Markov model of the dynamics is obtained using methods in Refs. [10,16]. The first nontrivial eigenvector ϕ_2 of the inferred Markov chain captures the long-time dynamics of the system, as discussed in the body of the paper. We represent the high-dimensional state space X_{K^*} through a two-dimensional (2D) UMAP embedding as in Ref. [16] (left), and color-code each point by its projection along ϕ_2 . An example 10-min-long centroid trajectory color-coded by ϕ_2 is shown on the right. The example showcases how negative and positive values of ϕ_2 correspond to forward “runs” and combinations of reversals, ventral and dorsal turns during “pirouettes.” (c) Example time series of ϕ_2 illustrating the stochastic hopping between runs and pirouettes. (d) Left: Distribution function of observing a run or a pirouette with a duration longer than τ_{beh} , $1 - P(\tau_{\text{beh}} \leq t)$, estimated from the experimental data (black), simulations of Eq. (2), i.e., the dynamics projected onto ϕ_2 (blue), and simulations of Eq. (1), i.e., of the full unprojected model (gray). While simulations capture the sum of exponential functions (gray dashed line) that approximates the bulk of the distribution, heavy tails observed in the data are not well captured. Right: Connected autocorrelation function $C_{\phi_2}(\tau)$ for the data (black) and simulations of the projected (unprojected) model [blue (gray)]. Simulations fail again in predicting the long-range correlations exhibited by the data. Note that the projected and the full model yield similar results, illustrating the efficiency of our projection method. Error bars represent 95% confidence intervals bootstrapped across worms.

the behavior into “runs” and “pirouettes” [16], as illustrated in Fig. 2(b). Large positive and negative values of ϕ_2 correspond to pirouettes and runs, respectively. This property allows us to define a slow reaction coordinate [32] that captures the worm’s dynamics along a “run-and-pirouette” axis [see Fig. 2(b)]. In particular, as the worm moves, it traces an orbit in the X_{K^*} space that we project onto ϕ_2 . The bottom line is that the fast dynamics of the body postures [9] is integrated out and the projection onto ϕ_2 highlights the effective stochastic description for the hopping between runs and pirouettes. A typical time series of $\phi_2(t)$ is shown in Fig. 2(c).

A. Inferring a stationary Langevin equation for the “run-and-pirouette” dynamics

To infer an explicit model for stochastic hoppings along $\phi_2(t)$, we sample the dynamics at the Markov-Einstein timescale τ^* [33,34], i.e., long enough that effects of higher-order eigenmodes ϕ_i ($i \geq 3$) have decayed. Thus, we can obtain an effective overdamped Langevin description for

$\phi_2(t)$ [35]:

$$\dot{\phi}_2 = F(\phi_2) + \sqrt{2D(\phi_2)}\eta(t), \quad (2)$$

where we effectively have $\langle \eta(t)\eta(t') \rangle \simeq \delta(t - t')$ due to the coarse sampling every τ^* . In practice, the choice $\tau^* = 0.75$ s ensures that a stochastic model inferred from the *C. elegans* time series results in effectively delta-correlated fluctuations (see Fig. S1(b) of the Supplemental Material [36]). To find $F(\phi_2)$ and $D(\phi_2)$ we use a kernel-based approach [37] based on the Kramers-Moyal expansion [38], rather than discretized bins, to obtain a more robust estimate (see Appendix A).

To probe the relevance of the above model, we identify “run” and “pirouette” states by maximizing the metastability of both states (see Appendix A) [16], and estimate the probability $P(\tau_{\text{beh}})$ of a residence time τ_{beh} in one of the two behaviors (see Fig. 2(d), left, and Fig. S2 [36]). Interestingly, while the exponential bulk of $P(\tau_{\text{beh}})$ is captured by Eq. (2), heavier tails are not. In addition, we estimated the connected

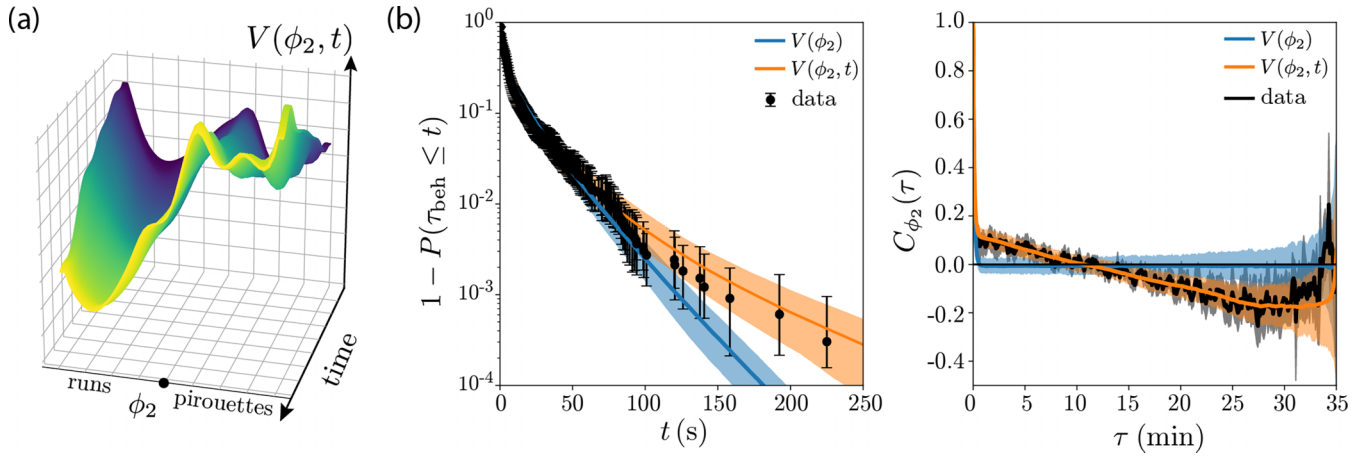


FIG. 3. A time-varying potential landscape captures heavy tails in *C. elegans* behavior. (a) The time-dependent potential landscape for the eigenmode ϕ_2 discussed in the text. As time goes on (blue to yellow), the barrier remains close to $\phi_2 = 0$ (black dot on the ϕ_2 axis) while the “run” well becomes deeper over time. That witnesses adaptation towards increasingly performing runs. (b) The probability of observing a run or a “pirouette” with a duration $> \tau_{\text{beh}}$, $1 - P(\tau_{\text{beh}} \leq t)$ (left), and the connected autocorrelation function, $C_{\phi_2}(\tau)$ (right), as obtained from data (black), simulations of the static model in Eq. (2) [blue, same as Fig. 1(d)], and simulations of the time-dependent model in Eq. (4) (orange). Note that the last captures heavy tails and long-range correlations observed in the data. Error bars represent 95% confidence intervals bootstrapped across worms.

autocorrelation function

$$C_{\phi_2}(\tau) = \frac{1}{\sigma_{\phi_2}^2} \langle (\phi_2(t) - \langle \phi_2 \rangle_t)(\phi_2(t + \tau) - \langle \phi_2 \rangle_t) \rangle_t, \quad (3)$$

where $\sigma_{\phi_2}^2$ is the variance of $\phi_2(t)$ and $\langle \cdot \rangle_t$ denotes temporal average. We observe again that the model captures short timescales (≈ 10 s) but fails to predict long-range correlations exhibited by the data [Fig. 2(d), right]. This discrepancy is not due to the projection onto ϕ_2 or the assumption of Langevin dynamics since simulations of the full model [Eq. (1)] yield similar predictions [see Fig. 2(d)].

B. Fluctuating potential landscapes underlie the emergence of heavy tails in *C. elegans* foraging

A possible explanation for the inability of Langevin equation (2) [or the full model in Eq. (1)] to capture heavy tails is the existence of subtle *hidden* fluctuations that evolve on timescales comparable to the observation time T_{expt} . The idea stems from observations that worms slowly adapt their search strategy upon removal of food by lowering their rate of “pirouettes” to explore wider areas in search for food [39–44]. A time-evolving rate of pirouettes calls for a nonstationary extension of the model via time-dependent drift and diffusion terms,

$$\dot{\phi}_2 = F(\phi_2, t) + \sqrt{2D(\phi_2, t)}\eta(t), \quad (4)$$

which reflect adaptation throughout T_{expt} .

Time-dependent drift and diffusion coefficients are inferred as described in Appendix A. The resulting potential landscape evolves as shown in Fig. 3(a), validating the hypothesis that worms slowly adapt by increasingly performing runs, in agreement with Refs. [39–44]. Over time, the “run-and-pirouette” random walk is indeed biased to explore farther away. Notably, time dependency is sufficient to reproduce

heavy tails and long-range correlations exhibited by the worms [Fig. 3(b)].

III. SLOWLY FLUCTUATING LANDSCAPES: GENERALITY OF HEAVY TAILS

Our data-driven results show that the observed heavy tails result from slow adaptation. Could similar mechanisms more generally underlie the widespread observation of heavy tails across behaving animals? Animals do modulate their behavior, either due to environmental factors or through endogenous internal states driven by neuromodulation, such as hunger or stress [45–47]. Such a continuum of scales inevitably results in nonstationarity since long-lived modes prevent relaxation within a finite observation time T_{expt} . Our goal here is to investigate theoretically the role of nonstationary fluctuations.

A. A fluctuating landscape picture of animal behavior

Given a set of observations of animal locomotion, we decompose the dynamics into ergodic, x , and nonergodic, s , components. The former are the variables that mix rapidly and define the potential wells that correspond to the stereotyped behaviors; the latter evolve on timescales $\tau_s \sim T_{\text{expt}}$ and slowly modulate the potential landscape of x . Assuming an appropriate timescale separation, we can describe the long-term dynamics by the following phenomenological model:

$$dx_t = -\tau_x^{-1} \partial_x U(x_t, s_t) dt + \sqrt{2T_x \tau_x^{-1}} dW_t^x, \quad (5)$$

$$ds_t = -\tau_s^{-1} \partial_s V(s_t) dt + \sqrt{2T_s \tau_s^{-1}} dW_t^s. \quad (6)$$

By rescaling time we can set $\tau_x = 1$; dW_t^x and dW_t^s are independent increments of a Wiener process, T_x and T_s are the levels of fluctuation in x and s , U is a potential landscape with multiple wells corresponding to long-lived stereotyped

behaviors, and V is uncoupled from the dynamics of x for simplicity.

In the following sections, we show that the slow modulation of the dynamics introduced by the hidden modes s generally give rise to heavy tails and nontrivial correlations, analogously to the above case of the worm. In addition, we determine the exponent of the tails and show that, in a limiting case, it asymptotes to the value -2 observed across animal species (see Fig. 1).

B. Heavy-tailed first-passage times in slowly driven metastable dynamics

In the context of the Langevin dynamics in Eq. (5), the distribution of times spent in a given behavioral state is given by the time to escape from a potential well, the so-called first-passage-time distribution (FPTD) [48], which is of interest across biology, chemistry, finance, physics, and mathematics [49–52]. We provide a short pedagogical introduction to first-passage times in Appendix B. Analytical expressions are rare [53] and most results focus on the mean first-passage time (MFPT) [54,55], more tractable but not representative of the long-time behavior in the presence of multiple timescales [56]. To investigate whether the nonergodic dynamics of Eqs. (5) and (6) generally yield heavy tails, we derive hereafter the large-time asymptotics of its FPTD.

The observation time T_{expt} separates ergodic and nonergodic modes and sets the slowest hopping rate $\omega_{\text{min}} \sim T_{\text{expt}}^{-1}$. The long-time behavior of the FPTD out of a static potential well is given by (see Appendix B)

$$f(t, \omega) = \omega e^{-\omega t}, \quad (7)$$

where the slowest ω_{min} normally dominates the asymptotic behavior. Since the typical time of modulation, $\tau_s = O(T_{\text{expt}})$, we can assume that transitions occur within a nearly static potential; i.e., we use Eq. (7) even in the presence of adaptation. However, ω_{min} will now vary as s fluctuates. The resulting FPTD $f(t)$ is given by the expectation value of $f(t, \omega)$ over the distribution $p(\omega)$ of ω weighted by the number of transitions within T_{expt} , which is $\propto \omega$. In short:

$$f(t) \propto \int_{\omega_{\text{min}}}^{\omega_{\text{max}}} p(\omega) \times \omega \times \omega e^{-\omega t} d\omega. \quad (8)$$

The tail of the distribution is dominated by instances where the barrier height is the largest, motivating the use of Kramers's approximation [48,57]:

$$\omega(s) = \omega_0 \exp \left\{ -\frac{\Delta U(s)}{T_x} \right\}, \quad (9)$$

where $\Delta U(s)$ is the height of the barrier to be overcome and ω_0 is a constant frequency (see Appendix B). Assuming that each measurement starts from initial conditions sampled according to a Boltzmann weight, the distribution of s is given by [58]

$$p(s) \propto \exp \left\{ -\frac{V(s)}{T_s} \right\}. \quad (10)$$

When the barrier height fluctuations are large enough to yield $\omega_{\text{min}}^{-1} \sim T_{\text{expt}}$, we can combine Eqs. (8)–(10) to obtain the

large- t limit of the FPTD:

$$f(t) \sim t^{-2} \exp \left\{ -\frac{V(\Delta U^{-1}(T_x \log(\omega_0 t)))}{T_s} \right\}. \quad (11)$$

Here, $\Delta U^{-1}(\cdot)$ represents the inverse function of $\Delta U(s)$ and we kept only the dominant order of the asymptotic approximation (see Appendix C). Importantly, when $T_s \rightarrow \infty$ we obtain $f(t) \sim t^{-2}$ under very general assumptions on the form of $V(s)$ and $U(x, s)$. In addition, when $V(s)$ and $\Delta U(s)$ are asymptotically equivalent, i.e., grow with the same power of s at large s , the distribution $f(t)$ behaves as a power law $f(t) \sim t^{-2-c\frac{T_x}{T_s}}$ with a correction to -2 proportional to $\frac{T_x}{T_s}$. In Fig. S3 [36] we confirm our theoretical predictions using numerical simulations of a Poisson process with varying hopping rates (see Appendix D). Equation (11) thus shows that slow modulation, which may result from interactions with the environment and/or slowly varying internal states [45,46], can indeed generally yield heavy-tailed FPTD.

Slowly driven double-well potential

As a further illustration of our result in Eq. (11), we consider a double-well potential whose barrier height is slowly modulated according to an Ornstein-Uhlenbeck process [Fig. 4(a)]. The dynamics of x and s are given by

$$\begin{aligned} dx_t &= -4s_t^2 x_t (x_t^2 - 1) dt + \sqrt{2T_x} dW_t^x, \\ ds_t &= -\tau_s^{-1} (s_t - \mu_s) dt + \sqrt{2T_s \tau_s^{-1}} dW_t^s, \end{aligned} \quad (12)$$

where $T_x = 10^{-3}$, $\mu_s = \sqrt{T_x}$, and $\tau_s = 10^3 T_{\text{expt}}$ (see Appendix A for details). Since the tail of $f(t)$ is dominated by large s values, we can take $V(s) \sim s^2/2$, and thus $V(\Delta U^{-1}(x)) \sim x/2$. Equation (11) predicts then

$$f(t) \sim t^{-2-\frac{T_x}{2T_s}}. \quad (13)$$

To test this result we performed direct numerical simulations of Eq. (12) while varying T_s and τ_s . Results in Figs. 4(b) and 4(c) (see also Fig. S4 [36]) quantitatively confirm the dependence of the power-law exponent on the ratio T_x/T_s , and its approaching t^{-2} as $T_s \rightarrow \infty$. These results support our theoretical predictions, and provide further intuition for how heavy-tailed distributions depend on the behavioral variability T_x , the strength of adaptation, T_s , and the timescale of behavioral adaptation, τ_s .

C. Long-range correlations and finite-size corrections in slowly driven metastable dynamics

This section complements the previous one by showing that slow modulation also induces heavy tails and long-range anticorrelations in the correlation function, as observed for the worm data in Fig. 3(b) (right).

1. Heavy tails

The connected correlation function of x in Eq. (5) is

$$C_x(\tau) = \frac{\langle x(t)x(t+\tau) \rangle - \langle x \rangle^2}{\langle x^2 \rangle - \langle x \rangle^2}, \quad (14)$$

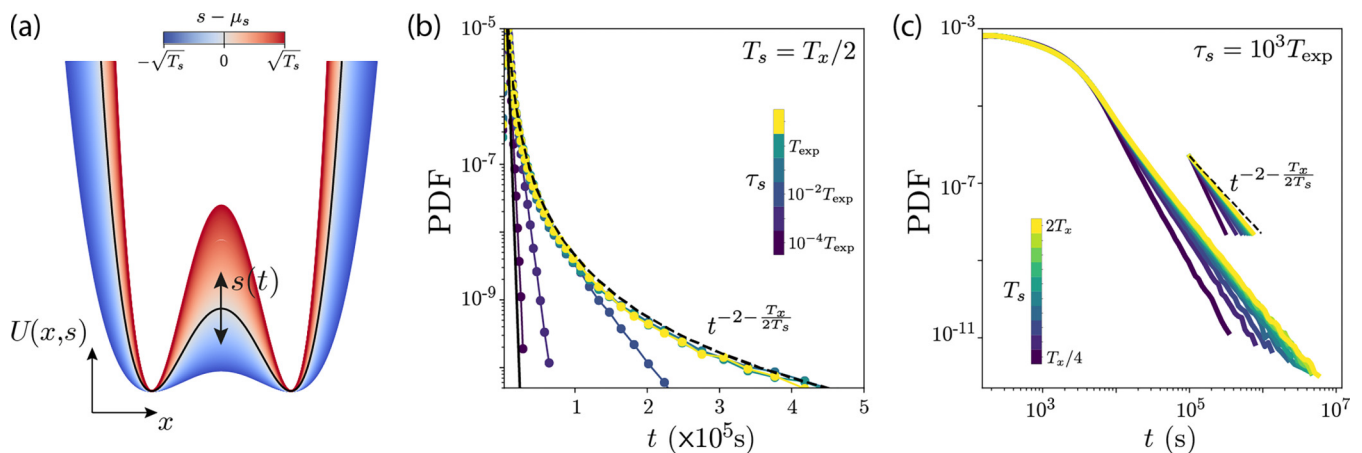


FIG. 4. Emergence of heavy tails in the first-passage-time distribution of a slowly driven double-well potential. (a) Schematic of the variation of the double-well potential with s (colored from blue to red; the black line represents $s = \mu_s$). (b) Probability density function (PDF) of times spent in a potential well (or equivalently, first passage time distribution) from numerical simulations of Eq. (12) for different values of τ_s and $T_s = T_x/2$ (see Appendix A). When $\tau_s \rightarrow 0$, the potential landscape relaxes to its mean value faster than the time to escape the well, resulting in an exponential with a hopping rate corresponding to μ_s (black line). As τ_s approaches T_{exp} , we observe a transition from exponential to power law, and in the limit of large τ_s we obtain the power law Eq. (13) (black dashed line). (c) PDF of times spent in a potential well from numerical simulations of Eq. (12) for large $\tau_s = 10^3 T_{\text{exp}}$ and different values of T_s (see Appendix A). As predicted, the tail of the distribution behaves as $f(t) \sim t^{-2-\frac{T_x}{2T_s}}$ (colored lines) with an exponent that approaches -2 as $T_s \rightarrow \infty$ (black dashed line).

where $\langle \cdot \rangle$ represents the ensemble average over the invariant density. In a static landscape, the long-time behavior of C_x is dominated by the first nontrivial eigenvalue of the Fokker-Planck operator Λ_2 , which is proportional to the slowest hopping rate $\Lambda_2 \propto \omega_{\min}$ [38,59], i.e., $C_x(\tau) \sim e^{-\Lambda_2 \tau}$. As in the previous section, when the landscape is slowly modulated, Λ_2 and ω_{\min} fluctuate, and C_x is given by a weighted average over $p(\omega)$:

$$C_x(\tau) \sim \int_{\omega_{\min}}^{\omega_{\max}} p(\omega) \times e^{-\omega \tau} d\omega. \quad (15)$$

Notice that, compared to Eq. (8), the integrand is divided by ω^2 : one ω is dropped since $C_x(t) \approx f(t, \omega)/\omega$ and the other ω is the number of hoppings, which ought to be counted for $f(t)$ but not for C_x . Following the same steps as for $f(t)$ (see Appendix E), we predict

$$C_x(\tau) \sim \exp \left\{ -\frac{V(\Delta U^{-1}(T_x \log(\omega_0 \tau)))}{T_s} \right\}, \quad (16)$$

to the dominant order for large τ 's. As for the FPTD, when $V(s)$ and $\Delta U(s)$ are asymptotically equivalent, $C_x(\tau) \sim$

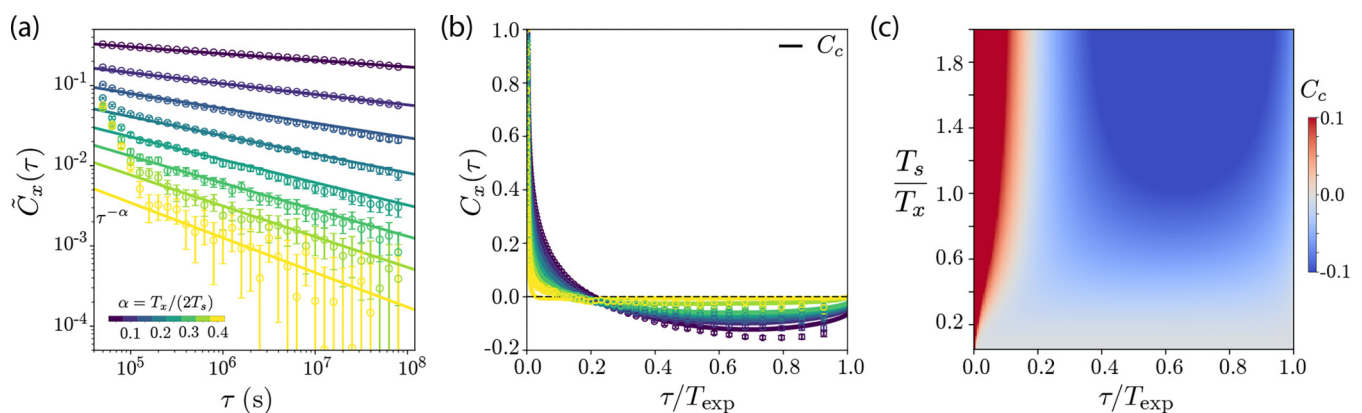


FIG. 5. Power-law correlations and finite-size corrections in a slowly driven double-well potential. (a) Estimated nonconnected correlation function $\tilde{C}_x(\tau) = \langle x(t)x(t+\tau) \rangle$ (see Appendix A) for the position x of a particle in a double-well potential driven on a timescale $\tau_s = 10^2 T_{\text{exp}}$. Our prediction $C_x(\tau) \sim \tau^{-\frac{T_x}{2T_s}}$ is validated (solid lines). Error bars represent 95% confidence intervals across 50 000 simulations. (b) Connected autocorrelation function, $C_x(\tau)$, directly estimated through time averages (see Appendix A), for a double-well potential driven on a timescale $\tau_s = 10^2 T_{\text{exp}}$. Due to the existence of timescales $\sim T_{\text{exp}}$, finite-size corrections C_c are present and generate long-range anticorrelations in $C_x(\tau)$, as predicted in Appendix F (solid lines). For both $C_x(\tau)$ and $C_c(\tau)$ we normalize the correlation functions by dividing by their value at $\tau = 1 \text{ lag} = 5 \times 10^{-4} T_{\text{exp}}$. Error bars represent 95% confidence intervals across 50 000 simulations. (c) Finite-size correction $C_c(\tau)$ vs T_s (see Appendix F). As we increase T_s , the range of observed ω grows, and so do finite-size corrections, which result in stronger anticorrelations (blue). Conversely, for small T_s finite-size effects are negligible as the longest sampled $\omega^{-1} \ll T_{\text{exp}}$.

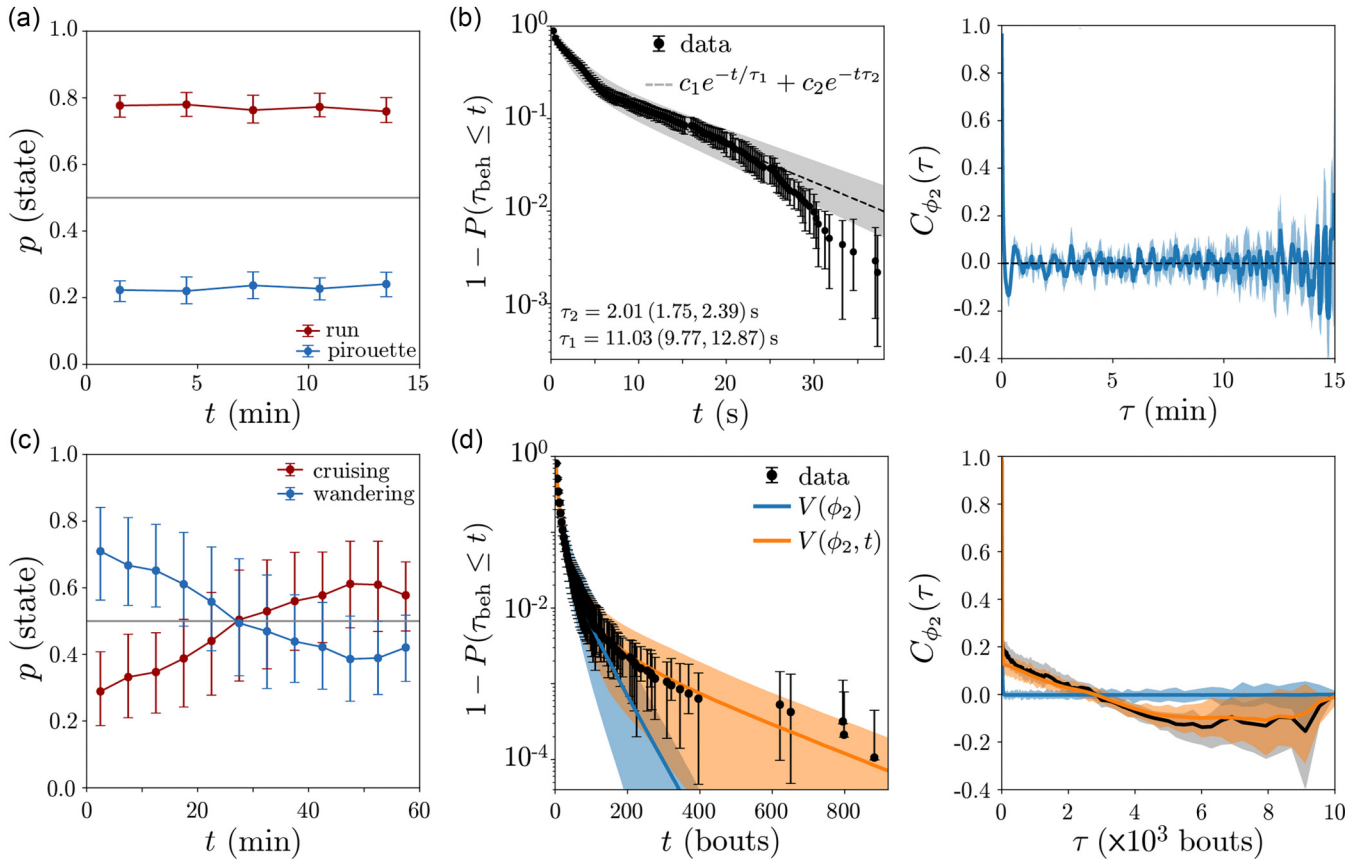


FIG. 6. Testing our theoretical predictions: loss-of-function mutation in the *npr-1* gene of *C. elegans* ablates heavy-tailed statistics, and larval zebrafish exhibit heavy tails in their long-lived behavior. (a) Probability of observing “run” and “pirouette” states as a function of time for *npr-1* mutants. Unlike wild-type N2 worms, the distribution of these states is constant over time. We estimate the fraction of time spent either performing a run or a pirouette in 3-min-long windows, and obtained bootstrapped averages across all seven worms. (b) Left: Complementary cumulative distribution function of observing a run or a pirouette with a duration $> \tau_{\text{beh}}$, $1 - P(\tau_{\text{beh}} \leq t)$ [same as Fig. 2(d), left]. The distribution is well approximated by a sum of exponential functions: a longer one that corresponds to the runs and a shorter one that corresponds to the pirouettes. Unlike N2 worms, the first-passage-time distribution in *npr-1* mutants is not heavy tailed. Right: Autocorrelation function C_{ϕ_2} of the slow mode ϕ_2 inferred from the *npr-1* mutant data. Unlike wild-type N2 worms, nontrivial long-range correlations are absent and instead correlations decay within a minute. (c) We collected data from Ref. [63] in which larval zebrafish are exposed to a chasing dot stimulus for 5 s every 2 min for at least 1 hour. We proceed as for *C. elegans* (see Appendix A) and find that the longest-lived dynamics ϕ_2 corresponds to a “wandering-cruising” axis, in which the fish either engages in bout sequences with large orientation changes (“wandering”), or performs sequences of smoother forward bouts (“cruising”), Fig. S7(b) [36]. Over time, we observe that the probability of the “cruising” state slowly increases over time, becoming more prevalent than the “wandering” state. We estimate the probability of being in the cruising and wandering states in 5-min windows, and show the average probability bootstrapped across the 11 fish. (d) Left: Complementary cumulative distribution function of observing cruising or wandering with a duration $> \tau_{\text{beh}}$, $1 - P(\tau_{\text{beh}} \leq t)$ [same as Fig. 2(d), left]. We plot the data in black, as well as the distribution obtained from simulations with a stationary model ($V(\phi_2)$, blue) and a time-dependent model ($V(\phi_2, t)$, orange) (see Appendix A for details). As for *C. elegans* N2 worms foraging, we find that time-dependent parameters are required to accurately predict the tail of the distribution. Right: Autocorrelation function C_{ϕ_2} of the slow mode dynamics obtained from the larval zebrafish data (black) as well as simulations from the stationary model (blue) and the time-dependent model (orange). Notably, we find that larval zebrafish exhibit heavy-tailed statistics due to the explicit time dependency of the behavioral dynamics.

$\tau^{-c \frac{\tau}{\tau_s}}$, with the same constant c . In this case, $f(t) \sim t^\beta$ and $C_x(\tau) \sim \tau^\gamma$ with exponents related by $\gamma = \beta + 2$.

To illustrate these results, we return to the double-well potential in Eq. (12). The expectation would be $C_x(\tau) \sim \tau^{-\frac{\gamma}{2\tau_s}}$ and $\langle x \rangle = 0$ (since the potential is symmetric). Indeed, if we measure the nonconnected correlation function from numerical simulations of Eq. (12) (without subtracting the mean; see Appendix A), we recover the theoretical expectation of power-law correlations for large τ_s , Fig. 5(a).

2. Finite-size effects and anticorrelations

In the previous double-well example, we used symmetry of the potential to avoid estimating the means. In general, we have to resort to empirical estimations via averaging in time (see Appendix A). In the presence of slow timescales, we expect that temporal averages $\hat{\mu}_x = \frac{1}{T_{\text{expt}}} \int_0^{T_{\text{expt}}} x(t) dt$ will deviate from ensemble averages, resulting in finite-size effects [60]. In particular, we expect that on average $x(t) - \hat{\mu}_x$ will change sign as time progresses. This transient behavior results in apparent long-range anticorrelations, since $x(t) - \hat{\mu}_x$ and

$x(t + \tau) - \hat{\mu}_x$ will tend to have different signs for large τ [61]. Therefore, we expect that Eq. (16) will deviate from empirical estimations when slow timescales are present. Indeed, when we estimate C_x through temporal averages (see Appendix A), we do observe long-range anticorrelations [see Fig. 5(b)]. Importantly, using our analytical derivation of the nonconnected correlation $\hat{C}_x(\tau)$ and results in Ref. [62], we derive an expression for the finite-size correction $C_c(\tau)$ (see Appendix F) that captures the empirical estimate of $C_x(\tau)$, as shown in Fig. 5(b).

Our predicted dependence of finite-size effects on T_s is well confirmed by results in Fig. 5. Furthermore, small τ_s yields the expected exponential tails for the nonconnected correlation function, which become a power law when $\tau_s \sim T_{\text{expt}}$ (see Fig. S5(a), left [36]). Note that finite-size corrections are apparent even in the exponential regime as long as its timescale is $\sim T_{\text{expt}}$ (see Figs. S5(a), right, and Fig. S5(b) [36]). This corresponds to barriers sufficiently high to produce hopping rates comparable to T_{expt} .

To conclude, note that our predictions recapitulate and rationalize the phenomenology of the correlations that we presented in Fig. 3(b-right) for foraging worms.

IV. TESTING OUR THEORETICAL PREDICTIONS

Our predictions in the previous section were motivated by the particular example of the wild-type nematode *C. elegans* but they emerged to be quite general. To challenge this predicted generality, we decided to consider different animal examples: a genetic mutant of *C. elegans* and the larval zebrafish. Both turn out to agree with our predictions and validate the idea that behavioral plasticity is a minimal yet necessary ingredient for the emergence of heavy tails in animal behavior.

A. A mutation in the *C. elegans npr-1* gene suppresses heavy tails observed in wild type animals

The NPR-1 neuropeptide receptor is known to impact several *C. elegans* behaviors, viz., aerotaxis and food response [45]. We collected a public data set where worms of the *npr-1* loss-of-function strain *npr-1(ad609)* are allowed to freely explore an agar plate with a uniform food patch (see Appendix A) and used the same method as in Fig. 2 for wild-type worms. The upshot is that the short-time behavior of *npr-1* mutants on food is similar to wild-type N2 worms off food. The structure of the behavioral landscape is similar to the one found in wild type [16], with the dominating kinetics being the transitions between “runs” and “pirouettes” (see Fig. S7(a) [36]). However, mutants and wild types crucially differ at long timescales as mutants do not exhibit heavy tails [see Fig. 6(b)]. Instead, the tail of the first-passage time is close to an exponential, and correlations decay to zero within a minute. This is due to the mutants’ inability to adapt their pirouette rates over time [Fig. 6(a)], contrary to the modulation highlighted in Fig. 3(a) for the wild type.

B. Heavy tails in the behavior of larval zebrafish results from slow habituation

Zebrafish larvae move in discrete tail bursts, interspersed by periods of immobility. We leverage a data set previously

analyzed in Refs. [63,64], where larvae are exposed to a “chasing dot” stimulus for 5 s every 2 mins for at least 1 hour (see Appendix A for details). The data set consists of a sequence of bouts, in which the curvature of the tail of the fish is tracked at a high spatiotemporal resolution yielding a time series of cumulative tail angles for each bout (see Fig. S7(b) [36]).

From these bout sequences, we proceed as for the postures of *C. elegans*, and find that $K^* = 5$ bouts and $\tau^* = 3$ bouts yield an accurate description of the long-lived dynamics (see Appendix A for details). Using these parameters, we find that the dominating long-lived mode ϕ_2 captures transitions between sequences of smooth forward bouts (“cruising”) and sequences of sharp orientation changes (“wandering”), as shown in Fig. S7(b) [36]. Projecting the full dynamics onto ϕ_2 yields a time series along this “cruising-wandering” axis that fluctuates over time.

Over long timescales, we observe that fish modulate time spent in the cruising state [Fig. 6(c)], likely due to the habituation to the stimulus condition. That is confirmed by the observation that, as in *C. elegans*, the inferred potential landscape is slowly varying in time, and this slow modulation is essential to accurately predict the heavy-tailed FPTD and the nontrivial long-range correlations shown in Fig. 6(d).

V. DISCUSSION

The combination of theory, numerics, and experimental data analysis we used here provided evidence that the multiplicity of timescales inherent to animal behavior is sufficient to give rise to heavy-tailed first-passage times and long-range correlations. The phenomenon is demonstrated in Fig. 2(d), which shows first-passage-time distributions and correlation functions for the *C. elegans* nematode. We started from this example as high-resolution measurements of the animal pose are available, bridging from ~ 0.1 s chaotic posture dynamics to ~ 10 s stochastic hopping among “runs” and “pirouettes”.

To capture long-term effects, we combined ideas from reduced-order modeling [32,65–67] and stochastic model inference [34,37,68]. The resulting one-dimensional stochastic differential equation yields an overdamped description of a partially observed system with metastable dynamics. Our contribution here is mostly methodological: rather than assuming structure *a priori*, we aim for a coarse-grained simplified description and let data drive the process of building it. While already effective, each step in our analysis can likely be enhanced by modern tools from machine learning (see, e.g., Ref. [69]), an issue that we leave for future work.

The static version of our coarse-grained description is unable to capture heavy tails and long-range correlations displayed by experimental data. This failure is not the result of our approximation, as a full, yet still autonomous, model of the dynamics is also missing these effects. Both are useful to capture short-term properties, but some new ingredient is needed to capture long times. The idea that we pursued here is that nonergodic modulations, varying over timescales comparable to the observation time T_{expt} , are the key. In practice, this amounts to having an effective potential landscape slowly changing in time. Figure 3 demonstrated that long-term

modulation allows us to capture the missing effects exhibited by the worm. In addition, we find that the slow modulation reflects the adaptation of the worm's foraging strategy to searching for food farther away. In this way, we go beyond classical approaches to reduced-order modeling, recognizing the existence of nonergodic fluctuations and introducing a nonstationary model to encode them.

Our analysis makes strong predictions that can be tested experimentally. In particular, perturbing the neural mechanisms responsible for the adaptation of “pirouette” rates, such as dopaminergic and glutamatergic signaling [39], should alter the long-term features of behavior. We predict that, in the absence of adaptation, shorter timescale movements remain unaffected while the dwell times in the “run” and “pirouette” states would become exponentially distributed, rather than heavy tailed, and that the correlation function would simply decay exponentially to zero on fast timescales. Mutants of the NPR-1 neuropeptide receptor that we analyzed in Sec. IV are our first promising and positive step in that direction. New mutants and data would be important to further confirm our predictions. It would also be interesting to check whether wild strains of *C. elegans* exhibit a similar pattern as the one found here for the laboratory strain N2. Indeed, laboratory strains of *C. elegans* were grown in relatively poor conditions for multiple generations, when compared to much richer natural environments encountered by wild strains [70–72]. This has led to an evolutionary divergence between the laboratory strain N2 and wild strains, with N2 worms fixing several mutations that affect a variety of phenotypes [73–76]. Whether the observed heavy tails are also observed in wild strains, and how they are modulated depending on the natural habitat of different worm species, would provide further insight into the ecological and evolutionary significance of heavy tails. We expect that richer environments yield behavioral modulations on a wider range of timescales and the exponent of the heavy tails will reflect the natural habitat of different species [77]. The explicit formulas (11) and (16) that we derived here should enlighten the contributions stemming from variations in the potential landscape (encoded in T_s), fluctuations in each behavior (encoded in T_x), and timescale of adaptation (encoded in τ_s). As for our predictions on long-range correlations, recent work provided evidence for power-law correlations in fruit flies [78]. Their observations are consistent with our theoretical predictions, and we argue that they might stem from nonergodic internal states. Indeed, we expect there to be slow modes that evolve on timescales comparable to the 1-hour recordings used in Ref. [78] (see Refs. [79,80]). Future work will be needed to shed light on this issue.

Heavy-tailed distributions in the duration t of behaviors with an exponent $f(t) \approx t^{-2}$ are found across multiple species, from bacteria [21], termites [81], and rats [20] to marine animals [19,82], humans [83], and even fossil records [84] (see Fig. 1). Such observations have led researchers to hypothesize that Lévy flights yield optimal search strategies when the power-law exponent is -2 [22–26], although this view has been met with some controversy [85,86]. Here, we provided evidence in two distant model organisms, *C. elegans* and zebrafish, that such fat tails may simply emerge from modulation over time and thus be a by-product of the evolutionarily favorable ability to perform adaptive behavior [87].

Power laws are observed beyond behavior, from solar flares [88,89] to the brain [90], and the idea of variable barrier heights has appeared in different contexts (for a review, see, e.g., Ref. [91]). In disordered systems, averaging over an exponential distribution of barrier heights can give rise to a broad distribution of waiting times [92,93]. Note that, while this mechanism is qualitatively analogous to the one presented here, ours relies on the temporal (rather than spatial) variation of barrier heights. That results in distinct emergent behavior that depends directly on the measurement timescale T_{expt} (that sets the lowest hopping rate ω_{min}) and the magnitude of the nonergodic fluctuations. Time-dependent energy barriers have also been used in bacterial chemotaxis [94]. The analysis of Ref. [94] concerns a particular limit of our derivation, in which $T_s \rightarrow \infty$ and the distribution of hopping rates becomes uniform. Our derivation considered a more general dynamics, and predicted corrections to the power laws that go beyond the limits in Ref. [94]. Another proposal is the presence of multiplicative noise terms in the dynamics [95,96], and this notion has recently been used to explain the emergence of Lévy flights in the collective behavior of midge swarms [97]. Our Eqs. (5) and (6) do give rise to an effectively colored multiplicative noise term for the quasistationary behavioral dynamics but we go beyond by determining the dependency on the relationship between the correlation time of the colored noise, τ_s , and the measurement time T_{expt} , and between the additive and multiplicative noise terms. Finally, some of the arguments we have put forward have appeared in discussions of “criticality” [98–100]. That is the apparent tendency of some systems to sit between two qualitatively different “phases” (see, e.g., Refs. [101]), which makes them akin to critical systems in statistical mechanics [102,103]. Our derivations here apply to a wider range of model classes, using the framework of out-of-equilibrium statistical mechanics to explicitly connect the long-timescale emergent behavior with the underlying effective fluctuations. In addition, unlike other approaches [99,104], our framework does not require explicit external drives, but simply collective modes that evolve in a weakly nonergodic fashion.

On the theoretical side, to derive the analytical expressions (11) and (16) of correlation function and FPTD, we exploited a separation between microscopic dynamics and long-time behavior. Further work will be required when such separation and the quasiadiabatic approximation do not hold. For example, we find numerically that for intermediate values of $1 \ll \tau_s \ll T_{\text{expt}}$ and finite T_s , the FPTD behaves as a truncated power law with an exponent > -2 . In this regime, the barrier heights fluctuate significantly before the particle hops. Intuitively, we expect that if barrier-crossing events become uncorrelated, the ω factor in the FPTD that accounts for the number of hopping events drops out, resulting in an exponent -1 rather than -2 [105]. In the opposite limit, when $\tau_s \gg T_{\text{expt}}$, it is the distribution of initial conditions (which we took to be Boltzmann) that determines the emergent behavior. This assumption holds if we consider that behavioral “individuality” is equivalent to having a very slow mode $\tau_s \gg T_{\text{expt}}$. This would mean that different animals in a population of conspecifics will exhibit a degree of “individuality” that matches the steady-state distribution of such long-lived modes. Intriguingly, such a relationship between interindividual variability

and long-lived behavioral variability is observed in flies [106]. In this sense, when $\tau_s \gg T_{\text{expt}}$, our results are equivalent to explaining the emergence of heavy tails through interindividual variability [107]. If such variability differs from the Boltzmann assumption, heavy tails need to be corrected accordingly, following the steps of our derivation but with a modified $p(s)$.

In conclusion, we have used a physics approach to shed light on animal behavior, leveraging statistical mechanics as a framework for thinking about the effect of slowly varying modulation, either environmental or by internal states. Concurrently, observations from animal behavior inspired new physical results regarding the emergence of heavy tails in slowly driven potential landscapes, which are generally relevant to a wide range of fields in chemistry, biology, or finance (see, e.g., Refs. [49–52,55,108,109] and references therein).

Code for reproducing our results is publicly available [110]. Data can be found in Ref. [111].

ACKNOWLEDGMENTS

We thank Adrian van Kan, Stéphan Fauve, Federica Ferretti, Tosif Ahamed, Nicola Rigolli and Arghyadip Mukherjee for their comments. We also thank Antonia Groneberg for sharing the dataset and João Marques for providing insight on larval zebrafish behavior. This work was partially supported by the LabEx ENS-ICFP: ANR-10-LABX-0010/ANR-10-IDEX-0001-02 PSL* and by the NIH Grant No. 1RF1NS128865-01. A.C. also acknowledges useful discussions at the Aspen Center for Physics, which is supported by the National Science Foundation, Grant No. PHY-1607611. G.S. was supported by the European Union’s Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie Grant No. 813457.

APPENDIX A: METHODS

1. *C. elegans* foraging data set

We used a previously analyzed data set [28], in which young-adult N2-strain *C. elegans* were tracked at $f = 16$ Hz [27]. Worms were grown at 20°C under standard conditions [112]. Before imaging, worms were removed from bacteria-strewn agar plates using a platinum worm pick, and rinsed from *E. coli* by letting them swim for 1 min in Nematode Growth Medium buffer. They were then transferred to an assay plate (9-cm Petri dish) that contained a copper ring (5.1 cm inner diameter) pressed into the agar surface, preventing the worm from reaching the side of the plate. Recording started approximately 6 min after the transfer and lasted for $T_{\text{expt}} = 35$ min.

2. Data-driven reduced-order model of *C. elegans* foraging dynamics

Building upon previous work [9,10,16], we extract a slow reaction coordinate that captures transitions between “runs” and “pirouettes” from the posture dynamics of *C. elegans*. As in Ref. [27], we extract the centerline of the animal at every $\delta t = 1/16$ s, estimate 100 tangent angles along the body, θ_t , and project these angles onto an eigenworm representation [28] to obtain a five-dimensional 35-min-long time

series, $\theta_t \rightarrow \vec{a}_t \in \mathbb{R}^5$, $t \in \{\delta t, 2\delta t, \dots, 33600\delta t\}$, that accurately captures the changing postures of each of the 12 animals (a combined total of 403 200 frames). From such measurements, we perform a time-delay embedding [9,16] to include short-term memory into an expanded maximally predictive state $X_{K^*}(t) = \vec{a}_{t-K^*:t}$, where $\vec{a}_{t-K^*:t} = \{\vec{a}_{t-K^*}, \dots, \vec{a}_{t-1}\}$. The amount of time delays K^* used to reconstruct the state space is chosen so as to maximize predictive information [10,16]. In this way, all the dynamics that mix on a sufficiently fast timescale (compared to the measurement time T_{expt}) should be included in the state. We note that, as evidenced throughout the paper, slow modes that evolve of timescales comparable to the measurement time do not provide enough independent samples to be directly inferred. Such nonstationary dynamics cannot be accounted for by a stationary model, and are thus not recoverable with a delay embedding due to their statistical insignificance. As in Ref. [16], we choose $K^* = 11$ frames = 0.6875 s to maximize predictive information within the limits of the data. We then partition the state space into a large number of discrete symbols through k -means clustering, and choose the number of partitions, $N^* = 1000$, so as to preserve as much information as possible in the discretization, while avoiding finite-size effects (as in Ref. [16]). The outcome of the partitioning is a symbolic sequence, where each symbol s_i corresponds to a small region of the state space (a collection of similar posture sequences). We then build a Markov chain by counting transitions among state-space partitions separated by a timescale τ , $P_{ij}(\tau) = P(s_j(t+\tau)|s_i(t)) \approx e^{\mathcal{L}\tau}$ [10,16], effectively approximating the action of the Perron-Frobenius operator (see, e.g., Ref. [113]). The transition time $\tau^* = 0.75$ s was chosen so as to self-consistently capture the long-lived dynamics [10,16]. The eigenfunctions of the Perron-Frobenius operator, and its adjoint, the Koopman operator, capture global patterns of the dynamics that relax to the steady-state distribution on different timescales. In particular, the slowest eigenfunctions of these operators offer optimal reaction coordinates that capture the slow dynamics of the system [66,67,114]. The slowest left eigenvector of the reversibilized transition matrix ϕ_2 captures transitions among “runs” and “pirouettes” that *C. elegans* uses to forage [16]. We find the transition point ϕ_2^c between runs and pirouettes by maximizing the overall coherence of the metastable states [10,16], and recenter and rescale ϕ_2 to have $\phi_2^c = 0$ at the transition point and to have equally spaced values within $[-2, \phi_2^c = 0]$ and $[\phi_2^c = 0, 2]$ [32]. Finally, each symbol s_i assumes a particular value of ϕ_2 , and so we can translate the symbolic sequence into a stochastic time series $\phi_2(t)$ that captures transitions between runs and pirouettes [see Figs. 2(a)–2(c)].

3. *npr-1* data set

We used a publicly available data set collected by the Schafer laboratory [115], in which young adults from the *C. elegans npr-1(ad609)* mutant strain were imaged for 15 min in food-rich plates. Worms were grown at 22°C under standard conditions, and adult worms were transferred to a 3.5-cm-diam plate with a nearly circular bacterial food lawn consisting of 20 μL of OP50 (see Ref. [115] for experimental details). After waiting 30 min for habituation, worms are then

imaged for a total of 15 min. We analyzed the videos using WORMPOSE [3] to resolve ambiguous poses when the worm is coiled and obtained a continuous time series of postures. From the full set of 12 worms analyzed with WORMPOSE, we kept only those that were successfully tracked for at least 80% of the frames and for which we could unambiguously assign a dorsal-ventral axis (through the bias of the a_3 distribution as in Ref. [3]). This resulted in a subset of seven worms that are fully tracked at 20 Hz for a combined total of 126 000 frames (18 000 per worm). We analyzed these data in the same way as the N2 wild-type worms: we project the posture onto an “eigenworm” space obtaining $\bar{a}_t \in \mathbb{R}^5$, $t \in \{\delta t, 2\delta t, \dots, 18000\delta t\}$, and extract the long-lived dynamics ϕ_2 using $K^* = 0.4$ s, $N^* = \text{int}(10^{2.75}) = 562$ partitions and $\tau^* = 0.5$ s. These parameters are chosen using the same criteria as for wild-type N2 worms: the full analysis pipeline can be found in Ref. [116]. We made the time series of eigenworm coefficients for the *npr-1* worms publicly available [111].

4. Zebrafish (*Danio rerio*) data set

We use a data set that was previously analyzed in Ref. [64]. See Refs. [63,64] for the experimental details. In these data, wild-type Tübingen fish (6–7 days postfertilization) are exposed to a “chasing dot” stimulus for 5 s every 2 min for at least 1 hour. The chasing dot stimulus consists of a dark spot with 1-mm radius that starts 2 cm away from the fish and approaches it from one of four directions (left, right, forward, or back). Video tracking was performed using custom methods (see Ref. [63]) that extract kinematic parameters from each burst of tail motion (bout). In our data set, each bout is represented by eight cumulative tail angles (measured along nine points from the swim bladder to the tip of the tail) over time for 250 ms, which corresponds to 175 frames sampled at 700 Hz. We project each bout, which is effectively a point in a 175 frames \times 8 tail angles dimensional space, to a 20-dimensional space using principal component analysis, preserving virtually all the significant information in the bout space (when compared to a shuffle). To gather enough long-timescale data, we selected recordings with at least 7500 bouts, yielding a total of 11 fish recordings and a combined total of 107 260 bouts. We have made the time series of bouts publicly available [111], and the analysis pipeline can be found in Ref. [116]. To extract the slow mode dynamics, we proceeded as we did with the *C. elegans* data: we reconstruct the long-lived dynamics ϕ_2 using $K^* = 5$ bouts, $\tau^* = 3$ bouts, and $N^* = \text{int}(10^{2.75}) = 562$ partitions, determining the parameters using the same approach as for the other data sets [10,16].

5. Two-dimensional UMAP embedding of the reconstructed state space

We use the UMAP embedding [117] as a tool to visualize the maximally predictive states X_{K^*} of *C. elegans* posture dynamics [16]. In a nutshell, the UMAP algorithm searches for a low-dimensional representation of the data that preserves its topological structure. We use a publicly available implementation of the algorithm [118], using Chebyshev distances. For wild-type N2 worms we use `n_neighbors = 50` nearest

neighbors and `min_dist = 0.05` as the minimum distance, while for the *npr-1* mutant data we use `n_neighbors = 15` and `min_dist = 0.05`, reflecting the smaller data set size.

6. Stochastic model inference

The Kramers-Moyal expansion transforms the master equation for the dynamics of Eq. (2) into a Fokker-Planck equation,

$$\begin{aligned} \partial_t \rho &= \mathcal{L} \rho = \partial_{\phi_2} J(\rho, \phi_2) \\ &= -\partial_{\phi_2} [F(\phi_2) \rho] + \partial_{\phi_2}^2 [D(\phi_2) \rho], \end{aligned} \quad (\text{A1})$$

where $J(\rho, \phi_2)$ is the current, and

$$\begin{aligned} F(x) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle \phi_2(t + \tau) - \phi_2(t) | \phi_2(t) = x \rangle, \\ D(x) &= \lim_{\tau \rightarrow 0} \frac{1}{2\tau} \langle (\phi_2(t + \tau) - \phi_2(t))^2 | \phi_2(t) = x \rangle. \end{aligned}$$

We use this expansion to estimate $F(\phi_2)$ and $D(\phi_2)$. In practice, given a time series Y_t , we estimate the averages in the Kramers-Moyal expansion using a kernel approach [37],

$$\begin{aligned} F(y)_{\tau,h} &= \frac{1}{\tau} \left\langle \frac{K_h(y - Y_t)}{\langle K_h(y - Y_{t'}) \rangle_{t'}} (Y_{t+\tau} - Y_t) \right\rangle_t, \\ D(y)_{\tau,h} &= \frac{1}{2\tau} \left\langle \frac{K_h(y - Y_t)}{\langle K_h(y - Y_{t'}) \rangle_{t'}} (Y_{t+\tau} - Y_t)^2 \right\rangle_t, \end{aligned}$$

where $K_h(z) = h^{-1} \kappa(z/h)$ and κ is the Epanechnikov kernel [119,120],

$$\kappa(y) = \begin{cases} \frac{3}{4\sqrt{5}} (1 - \frac{y^2}{5}) & \text{if } y^2 < 5 \\ 0 & \text{if } y^2 > 5. \end{cases}$$

Importantly, the estimator has an explicit dependence on the time delay τ and the bandwidth h . First, as discussed in the main text, we choose τ long enough such that most of the temporal correlations in the noise have decayed to zero. For *C. elegans* N2 worms crawling off food, it has been shown that $\tau^* = 0.75$ s gives an accurate first-order Markov model of the worm dynamics [16], and accordingly we find that a stochastic model inferred with $\tau^* = 0.75$ s yields nearly delta-correlated noise (Fig. S1(b) [36]). For the zebrafish data set, we use the same criterion to choose $\tau^* = 3$ bouts. Given this time delay τ^* , we choose the bandwidth through the Δ algorithm introduced in Ref. [37]. In essence, for each bandwidth h we estimate $F_{\tau^*,h}$ and $D_{\tau^*,h}$ and generate simulations with the estimated $F_{\tau^*,h}$ and $D_{\tau^*,h}$. From such simulations, we then re-infer the drift and diffusion from the simulated time series, obtaining $\hat{F}_{\tau^*,h}$ and $\hat{D}_{\tau^*,h}$. Finally, we compare the re-inferred drift and diffusion to the ones estimated directly from the time series,

$$\xi(h) = \frac{\int |f_{\tau^*,h} - \hat{f}_{\tau^*,h}| \sqrt{\pi(y) \hat{\pi}(y)} dy}{\int \sqrt{\pi(y) \hat{\pi}(y)} dy}, \quad (\text{A2})$$

where f can be either F or D , $\pi(y)$ is the steady-state distribution obtained from $F_{\tau^*,h}$ and $D_{\tau^*,h}$, and $\hat{\pi}(y)$ is the one obtained from $\hat{F}_{\tau^*,h}$ and $\hat{D}_{\tau^*,h}$. We choose h^* as the first minimum of $\xi(h)$ [37], locally minimizing the difference between

original and reconstructed drift and diffusion coefficients and avoiding the trivial minimum that corresponds to $h \rightarrow \infty$ (which yields constant F and D). For the *C. elegans* wild-type data of Fig. 3, we estimated the change in $\xi(h)$ as a function of h [Fig. S1(a)], which reaches zero at around $h^* \approx 0.1$. We choose $h^* = 0.08$ to infer the time series of $\phi_2(t)$. For the zebrafish data set, we use the same criterion to choose $h^* = 0.05$.

7. Nonstationary stochastic model inference

We proceed as before, but now infer F_{τ^*, h^*}^w and D_{τ^*, h^*}^w in overlapping windows. For the *C. elegans* wild-type data set we choose windows of length 5 min sampled every 30 s, while for zebrafish the overlapping windows were defined using 1000 bouts sampled every 100 bouts. The window length was chosen long enough to allow for equilibration of the long-lived dynamics, but also short enough such that the steady-state distribution remains approximately constant.

8. Reconstructing an effective potential landscape

From the Fokker-Planck equation, Eq. (A1), with natural boundary conditions, $J(\rho, \phi_2) = 0$, we can obtain the steady-state solution $\rho = \pi$, satisfying $\partial_t \pi = 0$, as

$$\pi(\phi_2) = \exp \left[\int \frac{F(\phi_2) - \partial_{\phi_2} D(\phi_2)}{D(\phi_2)} d\phi_2 \right].$$

Writing the steady-state distribution as a Boltzmann factor [121],

$$\pi(\phi_2) \propto e^{-\beta V(\phi_2)},$$

with $\beta = 1$, we can identify an effective potential landscape $V(\phi_2)$,

$$V(\phi_2) = - \int \frac{F(\phi_2) - \partial_{\phi_2} D(\phi_2)}{D(\phi_2)} d\phi_2.$$

The same approach applies to the time-dependent stochastic model, where each window has its own local steady state, and the effective potential landscape is time dependent:

$$V(\phi_2, t) = - \int \frac{F(\phi_2, t) - \partial_{\phi_2} D(\phi_2, t)}{D(\phi_2, t)} d\phi_2.$$

9. Stochastic model simulations of ϕ_2

We simulate the dynamics using an Euler scheme with the same sampling time as the data $\delta t = 1/16$ s or $\delta t = 1$ bout for the *C. elegans* N2 worms and zebrafish data sets, respectively. For the nonautonomous model, we take F_{τ^*, h^*}^w and D_{τ^*, h^*}^w of the window with a center closest to the sampled time point.

10. Fine-scale Markov model simulations

As in Ref. [16], we simulate symbolic sequences by sampling the next state according to the condition probability distribution $P(s_j(t + \tau^*) | s_i(t))$, which is simply the i th row of $P_{ij}(\tau^*)$. From this symbolic sequence, we can then obtain a simulated time series of $\phi_2(t)$ sampled on a timescale τ^* .

11. Estimating the first-passage-time distributions

We estimate the time spent either performing either state (“runs” and “pirouettes” for *C. elegans* and “cruising” and “wandering” for zebrafish) by identifying segments where $\phi_2(t) < 0$ (runs or cruising) or $\phi_2(t) > 0$ (pirouettes or wandering). To remove short-time fluctuations we subsample the data and the simulated time series by $\tau^*/2$.

12. Empirical estimate of the connected autocorrelation function

We estimate the connected autocorrelation function from M time traces at each lag $\tau = l\delta t$ as

$$\hat{C}_x(l\delta t) = \frac{1}{M} \sum_{\alpha=1}^M \frac{\frac{1}{N-l} \sum_{i=1}^N x_{\alpha,i} x_{\alpha,i+l} - \left(\frac{1}{N} \sum_{i=1}^N x_{\alpha,i} \right)^2}{\frac{1}{N} \sum_{i=1}^N x_{\alpha,i}^2 - \left(\frac{1}{N} \sum_{i=1}^N x_{\alpha,i} \right)^2}, \quad (\text{A3})$$

where $x_{\alpha,i}$ is the i th frame of the α trace with length $N = T_{\text{expt}}/\delta t$.

13. Estimating the first-passage-time distribution of a Poisson process with varying hopping rates

We sample s according to the Boltzmann distribution $p(s) \propto \exp(-\frac{(s-\mu_s)^2}{2T_s})$, and convert it to a hopping rate $\omega(s)$ by numerically integrating the backward Kolmogorov equation, Eq. (B2). We then sample first-passage-time events according to Eq. (7), until reaching the measurement timescale T_{expt} . We repeat this process 50 000 times, and collect the statistics of waiting times to build a normalized histogram of first-passage times with logarithmic bins, which we show in Fig. S3 [36].

14. Estimating the first-passage-time distribution in the slowly driven double-well potential

We generate 10 000 simulations of Langevin dynamics of Eq. (12), through an Euler scheme with a sampling time of $\delta t = 10^{-3}$ s for $T_{\text{expt}} = 10^7$ s. We then vary τ_s in the range $[10^{-4} T_{\text{expt}}, 10^4 T_{\text{expt}}]$ and T_s in the range $[T_x/4, 2T_x]$, where $T_x = 10^{-3}$ and $\mu_s = \sqrt{T_x}$. The initial conditions $x(0)$ are sampled randomly either as $x(0) = 1$ and $x(0) = -1$ with equal probability and $s(0) \sim \mathcal{N}(\mu_s, \sqrt{T_s})$ is sampled according to the Boltzmann distribution. From the simulations of $x(t)$, we then estimate the first-passage-time distribution by first identifying all the segments, $[t_0, t_f]$, in which t_0 corresponds to the first time x returns to $x_0 = \pm 1$ after reaching $x_f = 0$, and t_f is the time first to reach $x_f = 0$ after t_0 . Finally, we build a normalized histogram of first-passage times with logarithmic bins, which we show in Figs. 4 and S4 [36].

15. Estimating the autocorrelation functions in the slowly driven double-well potential

We generate 50 000 simulations of Langevin dynamics of Eq. (12), through an Euler scheme with an initial sampling time of $dt = 0.2$ s that is downsampled to $\delta t = 100$ s for $T_{\text{expt}} = 10^8$ s, with $T_x = 10^{-2}$, $\mu_x = \sqrt{T_x}$, τ_s sampled in the range $[10^{-2} T_{\text{expt}}, 10^2 T_{\text{expt}}]$, and T_s sampled in the range $[1.25T_x, 10T_x]$. We then estimate the connected autocorrelation function from the simulations using Eq. (A3). The

nonconnected correlation function is estimated as

$$\hat{C}_x(l\delta t) = \frac{1}{M} \sum_{\alpha=1}^M \frac{1}{N-l} \sum_{i=1}^N x_{\alpha,i} x_{\alpha,i+l},$$

and then normalized by dividing by $\hat{C}_x(l=1)$. The finite-size corrections to the correlation function are detailed in Appendix F.

APPENDIX B: A PRIMER ON FIRST-PASSAGE TIMES

We here give a brief introduction to the concept of first-passage times, following Refs. [48,122–124]. Broadly speaking, the first-passage-time probability is the probability that a random walker *first* reaches a particular site at a specified time. This concept is broadly relevant in biology: from the kinetics of on and off states in membrane receptors, to the binding and unbinding of transcription factors to DNA, to the dynamics of conformation changes in proteins, to the firing of neurons, and so on.

We here focus on the particular case of escape from a metastable state, which corresponds to first-passage time to reach the transition point. These are rare events: there is a timescale separation between the relaxation dynamics within a metastable state and the time it takes to cross the transition point. The average rate of escape from a metastable state (or the mean first-passage time to reach the transition point x_f) is given by a constant ω . The distribution of the first-passage times then corresponds to the probability of not escaping at each infinitesimal time step until time t , multiplied by the probability of escaping at time t . The probability to not jump until time t is given by $(1 - \omega\delta t)^{t/\delta t}$, since at each δt the probability of not transitioning is $1 - \omega\delta t$. Taking the limit of infinitesimal time steps gives $\lim_{\delta t \rightarrow 0} (1 - \omega\delta t)^{t/\delta t} = e^{-\omega t}$. Multiplying this by the probability of escaping at time t , the first-passage-time distribution in this simple case is given by $f(t, \omega) = \omega e^{-\omega t}$.

To determine ω , we use the backwards Kolmogorov equation to find the mean first-passage time, or mean exit time, which is defined as the expectation value of the first time $x(t)$ leaves a domain D , conditioned on $x(t) = x_0 \in D$. Let $\rho(y, t|x_0)$ represent the probability distribution of an ensemble of particles not leaving the domain D at time t . It satisfies the Fokker-Planck equation with absorbing boundary conditions,

$$\frac{d\rho}{dt} = \mathcal{L}\rho, \quad \rho(y, 0|x_0) = \delta(y - x_0), \quad \rho|_{\partial D} = 0, \quad (\text{B1})$$

where \mathcal{L} is the Fokker-Planck operator of the forward Kolmogorov equation. The total survival probability of still being inside the domain D at time t is obtained by integrating over the domain D ,

$$S(x_0, t) = \int_D \rho(y, t|x_0) dy.$$

Given the absorbing boundary conditions at the domain boundaries, this probability is a decreasing function of time,

$$\frac{\partial S}{\partial t} = -f(x_0, t),$$

where $f(x_0, t)$ is the first-passage-time distribution. The mean first-passage time is the first moment of this distribution,

$$\begin{aligned} \bar{\tau}(x_0) &= \int_0^\infty f(x_0, t) t dt = \int_0^\infty S(x_0, t) dt \\ &= \int_0^\infty \int_D \rho(y, t|x_0) dy dt. \end{aligned}$$

Writing the solution to Eq. (B1) as $\rho(y, t|x_0) = e^{\mathcal{L}t} \delta(y - x_0)$, we get

$$\begin{aligned} \bar{\tau}(x_0) &= \int_0^\infty \int_D e^{\mathcal{L}t} \delta(y - x_0) dy dt \\ &= \int_0^\infty (e^{\mathcal{L}t} 1)(x_0) dt. \end{aligned}$$

Applying the adjoint operator \mathcal{L}^* to the above equation, we obtain

$$\begin{aligned} \mathcal{L}^* \bar{\tau}(x_0) &= \int_0^\infty (\mathcal{L}^* e^{\mathcal{L}t} 1) dt = \int_0^\infty \frac{d}{dt} (e^{\mathcal{L}t} 1) dt \\ \mathcal{L}^* \bar{\tau}(x_0) &= -1. \end{aligned}$$

For a simple concrete example, consider the dynamics of Eq. (5) with $s_f = s \in \mathbb{R}$ constant. In this case, we can use the corresponding Fokker-Planck operator [see Eqs. (A1), (2), and (5)] to obtain

$$\begin{aligned} \mathcal{L}^* \bar{\tau}(x_0) &= -\tau_x^{-1} \partial_x U(x) \partial_x \bar{\tau}(x_0) + T_x \tau_x^{-1} \partial_x^2 \bar{\tau}(x_0) = -1, \\ \partial_x (e^{-U(x)/T_x} \partial_x \bar{\tau}) &= -\frac{e^{-U(x)/T_x}}{\tau_x T_x}. \end{aligned}$$

Let us consider that the particle starts at a minimum of the potential well $x_0 = a$ and escapes when reaching the barrier situated at $x_f = b$. Assuming reflective boundary conditions at $x_0 = a$ and an absorbing boundary at $x_f = b$ (where the escape events occurs), we have

$$\bar{\tau}(x_0) = \frac{1}{\tau_x T_x} \int_x^b e^{U(y)/T_x} dy \int_a^y e^{-\beta U(z)/T_x} dz.$$

Once the particle has reached $x = b$, there is a 50% chance that it manages to escape, yielding an escape rate

$$\omega^{-1} = \frac{2}{\tau_x T_x} \int_x^b e^{U(y)/T_x} dy \int_a^y e^{-\beta U(z)/T_x} dz. \quad (\text{B2})$$

When the barrier height is large when compared to the level of fluctuations, $\Delta U = U(x = b) - U(x = a) \gg T_x$, we can solve these integrals asymptotically to find

$$\omega = \frac{\omega_a \omega_b}{2\pi} e^{-\Delta U/T_x} = \omega_0 e^{-\Delta U/T_x},$$

where ω_a and ω_b correspond to the absolute value of the second derivative of the potential at $x = a$ and $x = b$ and stem from a Taylor expansion around the minimum and the maximum of the potential, respectively. We have thus recovered Kramers's approximation [48,57], Eq. (9).

APPENDIX C: FIRST-PASSAGE-TIME DISTRIBUTION IN SLOWLY FLUCTUATING POTENTIAL LANDSCAPES

We here derive the expression for the first-passage-time distribution (FPTD) in a fluctuating potential landscape. As

discussed in the main text, we consider the adiabatic limit in which the FPTD can be approximated by

$$f(t) \propto \int_{\omega_{\min}}^{\omega_{\max}} p(\omega) \omega^2 e^{-\omega t} d\omega,$$

where $\omega = \omega_0 \exp\{-\Delta U(s)/T_x\} \Rightarrow s(\omega) = \Delta U^{-1}(-T_x \log(\omega/\omega_0))$, and ω_0 is a typical (fast) frequency of the hopping dynamics [125]. The distribution $p(\omega)$ obeys $p(\omega)d\omega = p(s)ds$, where $p(s) \propto \exp\{-V(s)/T_s\}$, and is thus given by

$$p(\omega) \propto \exp\left\{-\frac{V(s(\omega))}{T_s}\right\} \frac{T_x/\omega}{\partial_s \Delta U(s)}.$$

Plugging this into Eq. (8), we get

$$f(t) \propto \int_{\omega_{\min}}^{\omega_{\max}} \exp\left\{-\frac{V(s(\omega))}{T_s}\right\} \frac{T_x}{\partial_s \Delta U(s)} \omega e^{-\omega t} d\omega. \quad (C1)$$

The exponential factor $e^{-\omega t}$ restricts the contributions to $\omega \sim 1/t$, which motivates the change of variable $\omega = \frac{\theta}{t}$. The above integral is then recast in the form

$$f(t) \propto t^{-2} \int_{\theta_{\min}(t)}^{\theta_{\max}(t)} \frac{\exp\left\{-\theta - \frac{V(s(\theta))}{T_s} + \log(\theta)\right\}}{\partial_s \Delta U(s(\theta))} d\theta, \quad (C2)$$

where $s(\theta) = \Delta U^{-1}(-T_x \log(\frac{\theta}{\omega_0 t}))$, $\theta_{\min}(t) = \omega_{\min} t$, and $\theta_{\max}(t) = \omega_{\max} t$.

To grasp the structure of the integral, it is convenient to consider first the special case where V and ΔU can be written as a power-series expansion $V(s) \sim as^n$ and $\Delta U(s) \sim bs^k$, $a, b \in \mathbb{R}$, with an equal dominant (at large values of the argument, see below) exponent n . The integral reduces then to the form

$$f(t) \propto t^{-2 - \frac{aT_x}{bT_s}} \int_{\theta_{\min}}^{\theta_{\max}} \frac{\theta^{1 + \frac{aT_x}{bT_s}} e^{-\theta}}{\left(-\log\left(\frac{\theta}{\omega_0 t}\right)\right)^{1 - \frac{1}{n}}} d\theta.$$

It remains to verify that the time dependencies at the denominator of the integrand and the limits of integration do not spoil the behavior at large times. This is verified by noting that the numerator of the integrand has the structure of an Euler Γ function of order $2 + \frac{aT_x}{bT_s}$. The denominator of the integrand has its maximum at $\theta^* = 1 + \frac{aT_x}{bT_s}$, decays over a range of values of order unity (we consider $\frac{aT_x}{bT_s}$ to be small or of order unity), and vanishes at the origin. In that range, the argument of the power at the denominator $\log(\omega_0 t) - \log(\theta) \simeq \log(\omega_0 t)$, which yields the final scaling with subdominant logarithmic corrections,

$$f(t) \sim t^{-2 - \frac{aT_x}{bT_s}} \times \log(\omega_0 t)^{\frac{1}{n} - 1}. \quad (C3)$$

To complete the argument, we note that the time dependency of θ_{\min} is not an issue as long as values $\theta \sim O(1)$ are in the integration range. In practice, this means that the minimum hopping rate ω_{\min} should be comparable to (or larger than) the measurement time, $\omega_{\min}^{-1} \sim O(T_{\text{expt}})$.

Before moving to the general case, two remarks are in order. First, for $\omega_0 t \gg 1$ the functions V and $\partial_s \Delta U$ that appear in Eq. (C2) have their argument $s \gg 1$. The dominant behavior of the two functions should then be understood for large values of their arguments. Second, the denominator $\partial_s \Delta U$ could *a priori* be included in the exponential at the numerator

but this does not modify our conclusion. It is indeed easy to verify that the maximum θ^* and the decay range would not be shifted at the dominant order (and this holds also for the general case considered hereafter).

We can now consider the general case with different dominant exponents $V(s) \sim as^n$ and $\Delta U(s) \sim bs^k$, $a, b \in \mathbb{R}$. The argument of the exponential in Eq. (C2),

$$L(\theta) = -\theta - \frac{V(s(\theta))}{T_s} + \log(\theta), \quad (C4)$$

has its maximum at θ^* , defined by the implicit equation

$$\theta^* = 1 + \frac{T_x}{T_s} \frac{\partial_s V(s(\theta^*))}{\partial_s \Delta U(s(\theta^*))} = 1 + \frac{T_x}{T_s} \frac{an}{bk} s^{n-k},$$

where we have used

$$\partial_\theta V(s) = \partial_s V(s) \times \frac{ds(\theta)}{d\theta}, \quad \frac{ds(\theta)}{d\theta} = -\frac{T_x/\theta}{\partial_s \Delta U(s)}.$$

For $n < k$, the maximum $\theta^* \simeq 1$ (as $s \gg 1$) and the integrand decays in a range of order unity. Indeed, the dominant order of the derivatives $\partial^p L$ ($p \geq 2$) at $\theta = \theta^*$ coincides with those of $\log(\theta)$. It follows that $L(\theta) - L(\theta^*) \simeq \log(\theta/\theta^*) - (\theta - \theta^*)$. The resulting integral over θ is an Euler Γ function of order 2, which indeed forms at values $O(1)$. In that range, $s \sim (\frac{T_x}{b} \log(\omega_0 t))^{1/k}$ and the integral is then approximated by $\exp\{L(\theta^*)\}$, so $f(t)$ becomes

$$f(t) \sim t^{-2} \exp\left\{-\frac{a\left(\frac{T_x}{b} \log(\omega_0 t)\right)^{n/k}}{T_s}\right\}.$$

The factor at the denominator in Eq. (C2) is $O[\exp\{(1/k - 1) \log[\log(\omega_0 t)]\}]$ and thus of the same order as terms that we have discarded in our approximation so we neglect it as well. Since the integral over θ forms for values $O(1)$, the constraint on the minimum hopping rate is the same as for the $n = k$ case, i.e., $\omega_{\min}^{-1} \sim O(T_{\text{expt}})$.

For $n > k$, the maximum $\theta^* \sim (\log \omega_0 t)^{n/k-1}$, which is now large. The dominant order of the derivatives $\partial^p L$ ($p \geq 2$) at $\theta = \theta^*$ is given by $(-1)^{p-1} (p-1)! (\theta^*)^{-(p-1)}$; that is, they coincide with those of $\theta^* \log(\theta)$. It follows that $L(\theta) - L(\theta^*) \simeq \theta^* [\log(\theta/\theta^*) - (\theta - \theta^*)/\theta^*]$. The resulting integral over θ is an Euler Γ function of (large) argument $\theta^* + 1$: its value is approximated by the Stirling formula, which yields $\int (\theta/\theta^*)^{\theta^*} e^{-(\theta-\theta^*)} d\theta \simeq \sqrt{\theta^*}$. The $\sqrt{\theta^*}$ reflects the fact that the integral forms around the maximum at θ^* of the integrand over a range $\sqrt{\theta^*}$, which implies that the approximation $-\log(\frac{\theta}{\omega_0 t}) \simeq \log(\omega_0 t)$ still holds, as in the previous cases $n \leq k$. The $\sqrt{\theta^*}$, as well as the $\log(\omega_0 t)^{1/k-1}$ coming from the denominator in Eq. (C2), are subdominant with respect to terms that we have neglected in the expansion of L . We therefore discard them from our final approximation for $n > k$:

$$f(t) \sim t^{-2} \exp\left\{-\frac{a\left(\frac{T_x}{b} \log(\omega_0 t)\right)^{n/k}}{T_s}\right\}.$$

Since the integral over θ forms for values $O((\log \omega_0 t)^{n/k-1}) \gg 1$, the condition $\omega_{\min}^{-1} \sim O(T_{\text{expt}})$ ensures *a fortiori* that the finite value of ω_{\min} does not affect the above result.

Discarding subdominant terms, in all three cases we thus get the general expression we present in the main text,

$$f(t) \sim t^{-2} \exp \left\{ -\frac{a \left(\frac{T_x}{b} \log(\omega_0 t) \right)^{n/k}}{T_s} \right\}. \quad (\text{C5})$$

To verify the validity of the above arguments, we show in Fig. S6 [36] how, to the dominant order, asymptotic predictions agree with a detailed numerical integration of Eq. (C1) for $\Delta U(s) = s^k$ and $V(s) = s^n$.

APPENDIX D: FIRST-PASSAGE TIME FOR A POISSON PROCESS WITH VARYING HOPPING RATES

To probe our theoretical predictions, we first assume that hopping events are well captured by a Poisson process and that the modulation of the potential landscape is infinitely slow such that the adiabatic approximation of Eq. (8) holds exactly. In practice, we sample s according to the Boltzmann distribution, Eq. (10), and, in order to relax from the Kramers approximation, we obtain $\omega(s)$ directly through numerically integrating the backward Kolmogorov equation, Eq. (B2) (with $\tau_x = 1$) [57].

We then sample events according to the distribution of first-passage times $f(t, \omega)$, Eq. (7), until reaching the measurement time T_{expt} [see Fig. S3(a) [36] for a schematic of the sampling procedure]. We take $U(x, s) = s^2(x^2 - 1)^2$ to be a symmetric double-well potential and sample s according to a Boltzmann distribution with $p(s) \propto \exp\{-\frac{(s-\mu_s)^2}{2T_s}\}$, corresponding to an Ornstein-Uhlenbeck process. Since the first-passage-time distribution is dominated by the large energy barriers, $s \gg \mu_s$, we take $V(\Delta U^{-1}(x)) \sim x/2$. From the derivation of Eq. (11), we expect that the final distribution of first-passage times will be given by

$$f(t) \sim t^{-2 - \frac{T_x}{2T_s}}. \quad (\text{D1})$$

Indeed, this is what we find through numerical simulations, Fig. S3(b) [36].

APPENDIX E: CORRELATION FUNCTIONS IN SLOWLY FLUCTUATING POTENTIAL LANDSCAPES

We here derive the expression for the correlation function in a fluctuating potential landscape. In general, the autocorrelation function can be expressed as a sum over exponential functions [38,59],

$$C_x(\tau) = \frac{\langle x(t)x(t+\tau) \rangle - \langle x \rangle^2}{\langle x^2 \rangle - \langle x \rangle^2} = \sum_i c_i e^{-\Lambda_i \tau},$$

where Λ_i ($\Lambda_1 < \Lambda_2 < \dots$) are the eigenvalues of the Fokker-Planck operator and $\sum_i c_i = 1$. Since we are interested in the long-term behavior of systems with energy barriers that can fluctuate over time, we assume that the large- τ behavior of the correlation function asymptotes to

$$C_x(\tau) \sim e^{-\Lambda_1 \tau},$$

where Λ_1 is the first nontrivial eigenvalue, which captures the longest-lived dynamics in the system. In addition, we assume that there is always a deeper well, with an escape rate ω , that

dominates the long-lived dynamics. In this case, $\Lambda_1 \propto \omega$ and we have

$$C_x(\tau) \sim e^{-\omega \tau}.$$

As previously discussed, we take the adiabatic approximation to derive the asymptotic behavior of the correlation function in the presence of slow nonergodic modulation of the potential landscape. In particular, we obtain a weighted average of the correlation function over multiple realizations of $\omega(s)$, yielding

$$C_x(\tau) \sim \int_{\omega_{\min}}^{\omega_{\max}} p(\omega) e^{-\omega \tau} d\omega. \quad (\text{E1})$$

Note that, in comparison with Eq. (8), besides dropping an ω factor due to the difference between $f(t, \omega)$, Eq. (7), and $C_x(\tau)$, we also do not need to take into account the extra factor of ω coming from the finite observation time, which in the case of the estimation of first-passage times biases the probability density in a manner that is proportional to ω . In the case of the correlation function, the dynamics of x is exposed to modulations in s , regardless of $\omega(s)$. Following the same steps as before, we consider that V and ΔU can be written as a series expansion with dominant terms $V(s) \sim as^n$ and $\Delta U(s) \sim bs^k$, $a, b \in \mathbb{R}$. In this case, we find that to dominant order

$$C_x(\tau) \sim \exp \left\{ -\frac{a \left(\frac{T_x}{b} \log(\omega_0 \tau) \right)^{n/k}}{T_s} \right\}. \quad (\text{E2})$$

Notably, when $n = k$, we obtain power-law correlations with an exponent that depends on the ratio of temperatures,

$$C_x(\tau) \sim \tau^{-\frac{\alpha T_x}{b T_s}} \times \log(\omega_0 \tau)^{\frac{1}{n}-1}, \quad (\text{E3})$$

where we have included subdominant corrections. In particular, we find that when $n = k$ both the first-passage-time distribution $f(t) \sim t^\beta$ and the correlation function $C_x(\tau) \sim \tau^\gamma$ have power-law behavior at large times, and that the exponents are related by $\gamma = \beta + 2$. In addition, when $T_s \rightarrow \infty$ correlations decay slowly as $C_x(\tau) \sim \log(\omega_0 \tau)^{1/n-1}$.

APPENDIX F: FINITE-SIZE CORRECTION TO THE CORRELATION FUNCTION

When estimating the correlation function from a collection of M finite time traces sampled at δt and with length $N = T_{\text{expt}}/\delta t$, we compute

$$\hat{C}_x(\tau = l\delta t) = \frac{1}{M} \sum_{\alpha=1}^M \frac{\frac{1}{N-l} \sum_{i=1}^N x_{\alpha,i} x_{\alpha,i+l} - \left(\frac{1}{N} \sum_{i=1}^N x_{\alpha,i} \right)^2}{\frac{1}{N} \sum_{i=1}^N x_{\alpha,i}^2 - \left(\frac{1}{N} \sum_{i=1}^N x_{\alpha,i} \right)^2},$$

where $x_{\alpha,i}$ is the i th frame of the α trace. Assuming that the finite-size corrections to the correlation function are dominated by corrections to the mean value (and not the variance), we can leverage the derivation of Desponds *et al.* [62] to obtain an expression for the finite-size corrections to the correlation function from the nonconnected correlation function

$$\tilde{C}_x = \langle x(t)x(t + \tau) \rangle,$$

$$C_c(\tau) \sim \tilde{C}(l) + \frac{1}{N} \left(\frac{1}{N} - \frac{2}{N-l} \right) \left(N\tilde{C}(0) + \sum_{k=1}^{N-1} 2(N-k)\tilde{C}(k) \right) + \frac{2}{N(N-l)} \left(l\tilde{C}(0) + \sum_{k=1}^{l-1} 2(l-k)\tilde{C}(k) + \sum_{m=1}^{N-1} \tilde{C}(m)(\min(m+l, N) - \max(l, m)) \right). \quad (\text{F1})$$

As detailed in the main text, as a case study we take the overdamped dynamics for the position x of a particle in a symmetric double-well potential, for which the barrier height can fluctuate according to a slow parameter s , Eq. (12). The timescale separation between the hopping events and the relaxation to the well means that the correlation function is dominated by the first nontrivial eigenvalue, $C_x(\tau) \sim e^{-\Lambda_1\tau} = e^{-2\omega\tau}$, where we take $\Lambda = 2\omega$ due to the fact that the potential wells have the same depth [38]. Taking $V(s) \sim s^2/2$ and $\Delta U(s) = s^2$ we obtain that, in the asymptotic large- τ limit, $C_x(\tau) \sim \tau^{-\frac{1}{2s}} \log(\omega_0\tau)^{-1/2}$, Fig. 5(a).

To obtain an accurate estimate of the correlation function for all τ , we go beyond the asymptotic approximation and

numerically integrate

$$\tilde{C}_x(\tau) \sim \int_{-\infty}^{\infty} e^{-V(s)/T_s} e^{-2\omega(s)\tau} ds, \quad (\text{F2})$$

where $\omega(s)$ can be estimated directly by integrating the Kolmogorov backward equation. At large τ , the numerical integration of $\tilde{C}_x(\tau)$ matches the asymptotic behavior $\tau^{-\frac{1}{2s}}$. Plugging Eq. (F2) into Eq. (F1), we obtain the correction to the autocorrelation function $C_c(\tau)$ presented in Figs. 5(b) and 5(c).

-
- [1] T. D. Pereira, J. W. Shaveitz, and M. Murthy, Quantifying behavior to understand the brain, *Nat. Neurosci.* **23**, 1537 (2020).
- [2] M. W. Mathis and A. Mathis, Deep learning tools for the measurement of animal behavior in neuroscience, *Curr. Opin. Neurobiol.* **60**, 1 (2020).
- [3] L. Hebert, T. Ahamed, A. C. Costa, L. O’Shaughnessy, and G. J. Stephens, WormPose: Image synthesis and convolutional networks for pose estimation in *C. elegans*, *PLoS Comput. Biol.* **17**, e1008914 (2021).
- [4] G. J. Berman, Measuring behavior across scales, *BMC Biol.* **16**, 23 (2018).
- [5] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters and Complexity*, 1st ed. (Oxford University Press, Oxford, UK, 2006).
- [6] J. Crank, *The Mathematics of Diffusion* (Clarendon Press, Oxford, UK, 1979).
- [7] S. Brenner, The genetics of *Caenorhabditis elegans*, *Genetics* **77**, 71 (1974).
- [8] C. I. Bargmann and E. Marder, From the connectome to brain function, *Nat. Methods* **10**, 483 (2013).
- [9] T. Ahamed, A. C. Costa, and G. J. Stephens, Capturing the continuous complexity of behaviour in *Caenorhabditis elegans*, *Nat. Phys.* **17**, 275 (2021).
- [10] A. C. Costa, T. Ahamed, D. Jordan, and G. J. Stephens, Maximally predictive states: From partial observations to long timescales, *Chaos* **33**, 023136 (2023).
- [11] F. Takens, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. Rand and L.-S. Young (Springer, Berlin, 1981), pp. 366–381.
- [12] G. Sugihara and R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature (London)* **344**, 734 (1990).
- [13] T. Sauer, J. A. Yorke, and M. Casdagli, Embedology, *J. Stat. Phys.* **65**, 579 (1991).
- [14] J. Stark, Delay embeddings for forced systems. I. Deterministic forcing, *J. Nonlinear Sci.* **9**, 255 (1999).
- [15] J. Stark, D. S. Broomhead, M. Davies, and J. Huke, Delay embeddings for forced systems. II. Stochastic forcing, *J. Nonlinear Sci.* **13**, 519 (2003).
- [16] A. C. Costa, T. Ahamed, D. Jordan, and G. J. Stephens, A Markovian dynamics for *C. elegans* behavior across scales, [arXiv:2310.12883](https://arxiv.org/abs/2310.12883).
- [17] J. T. Pierce-Shimomura, T. M. Morse, and S. R. Lockery, The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis, *J. Neurosci.* **19**, 9557 (1999).
- [18] H. C. Berg, *E. coli in Motion*, Biological and Medical Physics Series (Springer, New York, 2004).
- [19] D. W. Sims, E. J. Southall, N. E. Humphries, G. C. Hays, C. J. A. Bradshaw, J. W. Pitchford, A. James, M. Z. Ahmed, A. S. Brierley, M. A. Hindell, D. Morritt, M. K. Musyl, D. Righton, E. L. C. Shepard, V. J. Wearmouth, R. P. Wilson, M. J. Witt, and J. D. Metcalfe, Scaling laws of marine predator search behaviour, *Nature (London)* **451**, 1098 (2008).
- [20] K. Jung, H. Jang, J. D. Kralik, and J. Jeong, Bursts and heavy tails in temporal and sequential dynamics of foraging decisions, *PLoS Comput. Biol.* **10**, e1003759 (2014).
- [21] E. Korobkova, T. Emonet, J. M. G. Vilar, T. S. Shimizu, and P. Cluzel, From molecular noise to behavioural variability in a single bacterium, *Nature (London)* **428**, 574 (2004).
- [22] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. da Luz, E. P. Raposo, and H. E. Stanley, Optimizing the success of random searches, *Nature (London)* **401**, 911 (1999).
- [23] M. E. Wosniack, M. C. Santos, E. P. Raposo, G. M. Viswanathan, and M. G. E. da Luz, Robustness of optimal

- random searches in fragmented environments, *Phys. Rev. E* **91**, 052119 (2015).
- [24] M. E. Wosniack, M. C. Santos, E. P. Raposo, G. M. Viswanathan, and M. G. E. da Luz, The evolutionary origins of Levy walk foraging, *PLoS Comput. Biol.* **13**, e1005774 (2017).
- [25] B. Guinard and A. Korman, Intermittent inverse-square Levy walks are optimal for finding targets of all sizes, *Sci. Adv.* **7**, eabe8211 (2021).
- [26] A. Clementi, F. d'Amore, G. Giakkoupis, and E. Natale, Search via Parallel Lévy Walks on Z^2 , in *Proceedings of the 2021 ACM Symposium on Principles of Distributed Computing* (ACM Press, New York, 2021), pp. 81–91.
- [27] O. D. Broekmans, J. B. Rodgers, W. S. Ryu, and G. J. Stephens, Resolving coiled shapes reveals new reorientation behaviors in *C. elegans*, *eLife* **5**, e17227 (2016).
- [28] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, Dimensionality and dynamics in the behavior of *C. elegans*, *PLoS Comput. Biol.* **4**, e1000028 (2008).
- [29] H. Mori, Transport, collective motion, and brownian motion, *Prog. Theor. Phys.* **33**, 423 (1965).
- [30] R. Zwanzig, Nonlinear generalized Langevin equations, *J. Stat. Phys.* **9**, 215 (1973).
- [31] A. Rupe, V. V. Vesselinov, and J. P. Crutchfield, Nonequilibrium statistical mechanics and optimal prediction of partially-observed complex systems, *New J. Phys.* **24**, 103033 (2022).
- [32] G. Froyland, G. A. Gottwald, and A. Hammerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, *SIAM J. Appl. Dyn. Syst.* **13**, 1816 (2014).
- [33] R. Friedrich, J. Peinke, M. Sahimi, and M. Reza Rahimi Tabar, Approaching complexity by stochastic methods: From biological systems to turbulence, *Phys. Rep.* **506**, 87 (2011).
- [34] J. L. Callahan, J.-C. Loiseau, G. Rigas, and S. L. Brunton, Nonlinear stochastic modelling with Langevin regression, *Proc. R. Soc. A* **477**, 20210092 (2021).
- [35] We use the Itô interpretation of the stochastic dynamics (see, e.g., Ref. [126]).
- [36] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PRXLife.2.023001> for supplemental figures that support our main results.
- [37] D. Lamoignon and K. Lehnertz, Kernel-based regression of drift and diffusion coefficients of stochastic processes, *Phys. Lett. A* **373**, 3507 (2009).
- [38] H. Risken and H. Haken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 3rd ed. (Springer, Berlin, 1989).
- [39] T. Hills, P. J. Brockie, and A. V. Maricq, Dopamine and glutamate control area-restricted search behavior in *Caenorhabditis elegans*, *J. Neurosci.* **24**, 1217 (2004).
- [40] J. M. Gray, J. J. Hill, and C. I. Bargmann, A circuit for navigation in *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. USA* **102**, 3184 (2005).
- [41] L. C. Salvador, F. Bartumeus, S. A. Levin, and W. S. Ryu, Mechanistic analysis of the search behaviour of *Caenorhabditis elegans*, *J. R. Soc. Interface* **11**, 20131092 (2014).
- [42] A. J. Calhoun, A. Tong, N. Pokala, J. A. J. Fitzpatrick, T. O. Sharpee, and S. H. Chalasani, Neural mechanisms for evaluating environmental variability in *Caenorhabditis elegans*, *Neuron* **86**, 428 (2015).
- [43] I. Hums, J. Riedl, F. Mende, S. Kato, H. S. Kaplan, R. Latham, M. Sonntag, L. Traunmüller, and M. Zimmer, Regulation of two motor patterns enables the gradual adjustment of locomotion strategy in *Caenorhabditis elegans*, *eLife* **5**, e14116 (2016).
- [44] S. W. Flavell, D. M. Raizen, and Y.-J. You, Behavioral states, *Genetics* **216**, 315 (2020).
- [45] C. I. Bargmann, Beyond the connectome: How neuromodulators shape neural circuits, *BioEssays* **34**, 458 (2012).
- [46] E. Marder, Neuromodulation of neuronal circuits: Back to the future, *Neuron* **76**, 1 (2012).
- [47] S. W. Flavell, N. Gogolla, M. Lovett-Barron, and M. Zelikowsky, The emergence and influence of internal states, *Neuron* **110**, 2545 (2022).
- [48] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, *Rev. Mod. Phys.* **62**, 251 (1990).
- [49] A. Szabo, K. Schulten, and Z. Schulten, First passage time approach to diffusion controlled reactions, *J. Chem. Phys.* **72**, 4350 (1980).
- [50] S. Condamin, V. Tejedor, R. Voituriez, O. Bénichou, and J. Klafter, Probing microscopic origins of confined subdiffusion by first-passage observables, *Proc. Natl. Acad. Sci. USA* **105**, 5675 (2008).
- [51] O. Bénichou and R. Voituriez, From first-passage times of random walks in confinement to geometry-controlled kinetics, *Phys. Rep.* **539**, 225 (2014).
- [52] R. Chichportiche and J.-P. Bouchaud, in *First-Passage Phenomena and Their Applications* (World Scientific, Singapore, 2014), Chap. 1, pp. 447–476.
- [53] D. S. Grebenkov, First exit times of harmonically trapped particles: A didactic review, *J. Phys. A: Math. Theor.* **48**, 013001 (2015).
- [54] P. Hänggi, Escape over fluctuating barriers driven by colored noise, *Chem. Phys.* **180**, 157 (1994).
- [55] O. Bénichou, T. Guérin, and R. Voituriez, Mean first-passage times in confined media: From Markovian to non-Markovian processes, *J. Phys. A: Math. Theor.* **48**, 163001 (2015).
- [56] A. Godec and R. Metzler, Universal proximity effect in target search kinetics in the few-encounter limit, *Phys. Rev. X* **6**, 041037 (2016).
- [57] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [58] We note that this is generally true even for fixed initial conditions as long as $\tau_s < T_{\text{expt}}$. If $\tau_s \sim T_{\text{expt}}$, then $p(\omega)$ is primarily defined by the distribution of initial conditions. However, when the initial conditions are well approximated by a normal distribution with variance σ^2 , the denominator in the Boltzmann weight should be changed accordingly and this will change the final form of the first-passage-time distribution. Nonetheless, the derivation we present is general and can be adapted for a given $p(s)$ (see Appendix C).
- [59] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin Equation*, 2nd ed. (World Scientific, Singapore, 2004).
- [60] A. Cavagna, I. Giardina, and T. S. Grigera, The physics of flocking: Correlation as a compass from experiments to theory, *Phys. Rep.* **728**, 1 (2018).
- [61] A similar observation has been done in the analysis of spatial correlation functions in flocks of birds [127].
- [62] J. Desponds, H. Tran, T. Ferraro, T. Lucas, C. Perez Romero, A. Guillou, C. Fradin, M. Coppey, N. Dostatni, and A. M.

- Walczak, Precision of readout at the hunchback gene: Analyzing short transcription time traces in living fly embryos, *PLoS Comput. Biol.* **12**, e1005256 (2016).
- [63] J. C. Marques, S. Lackner, R. Félix, and M. B. Orger, Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering, *Curr. Biol.* **28**, 181 (2018).
- [64] A. H. Groneberg, J. C. Marques, A. L. Martins, R. Diez del Corral, G. G. de Polavieja, and M. B. Orger, Early-life social experience shapes social avoidance reactions in larval zebrafish, *Curr. Biol.* **30**, 4009 (2020).
- [65] D. Givon, R. Kupferman, and A. Stuart, Extracting macroscopic dynamics: Model problems and algorithms, *Nonlinearity* **17**, R55 (2004).
- [66] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems, *Multiscale Model. Simul.* **7**, 842 (2008).
- [67] D. Giannakis, Data-driven spectral decomposition and forecasting of ergodic dynamical systems, *Appl. Comput. Harmon. Anal.* **47**, 338 (2019).
- [68] A. Frishman and P. Ronceray, Learning force fields from stochastic trajectories, *Phys. Rev. X* **10**, 021009 (2020).
- [69] F. Dietrich, A. Makeev, G. Kevrekidis, N. Evangelou, T. Bertalan, S. Reich, and I. G. Kevrekidis, Learning effective stochastic differential equations from microscopic simulations: Linking stochastic numerics to deep learning, *Chaos* **33**, 023121 (2023).
- [70] M.-A. Félix and C. Braendle, The natural history of *Caenorhabditis elegans*, *Curr. Biol.* **20**, R965 (2010).
- [71] L. Frézal and M.-A. Félix, The natural history of model organisms: *C. elegans* outside the Petri dish, *eLife* **4**, e05849 (2015).
- [72] H. Schulenburg and M.-A. Félix, The natural biotic environment of *Caenorhabditis elegans*, *Genetics* **206**, 55 (2017).
- [73] P. T. McGrath, M. V. Rockman, M. Zimmer, H. Jang, E. Z. Macosko, L. Kruglyak, and C. I. Bargmann, Quantitative mapping of a digenic behavioral trait implicates Globin Variation in *C. elegans* sensory behaviors, *Neuron* **61**, 692 (2009).
- [74] J. W. M. Green, J. J. Stastna, H. E. Orbidans, and S. C. Harvey, Highly polygenic variation in environmental perception determines dauer larvae formation in growing populations of *Caenorhabditis elegans*, *PLoS ONE* **9**, e112830 (2014).
- [75] E. C. Andersen, J. S. Bloom, J. P. Gerke, and L. Kruglyak, A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology, *PLoS Genet.* **10**, e1004156 (2014).
- [76] M. G. Sterken, L. B. Snoek, J. E. Kammenga, and E. C. Andersen, The laboratory domestication of *Caenorhabditis elegans*, *Trends Genet.* **31**, 224 (2015).
- [77] I. Beets, G. Zhang, L. A. Fenk, C. Chen, G. M. Nelson, M.-A. Félix, and M. de Bono, Natural variation in a dendritic scaffold protein remodels experience-dependent plasticity by altering neuropeptide expression, *Neuron* **105**, 106 (2020).
- [78] W. Bialek and J. W. Shaevitz, Long timescales, individual differences, and scale invariance in animal behavior, *Phys. Rev. Lett.* **132**, 048401 (2024).
- [79] B. Qiao, C. Li, V. W. Allen, M. Shirasu-Hiza, and S. Syed, Automated analysis of long-term grooming behavior in *Drosophila* using a k-nearest neighbors classifier, *eLife* **7**, e34497 (2018).
- [80] K. E. Overman, D. M. Choi, K. Leung, J. W. Shaevitz, and G. J. Berman, Measuring the repertoire of age-related behavioral changes in *Drosophila melanogaster*, *PLoS Comput. Biol.* **18**, e1009867 (2022).
- [81] O. Miramontes, O. DeSouza, L. R. Paiva, A. Marins, and S. Orozco, Lévy flights and self-similar exploratory behaviour of termite workers: Beyond model fitting, *PLoS ONE* **9**, e111183 (2014).
- [82] N. E. Humphries, N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. R. Houghton, G. C. Hays, C. S. Jones, L. R. Noble, V. J. Wearmouth, E. J. Southall, and D. W. Sims, Environmental context explains Lévy and Brownian movement patterns of marine predators, *Nature (London)* **465**, 1066 (2010).
- [83] D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. P. Mabulla, F. W. Marlowe, and H. Pontzer, Evidence of Lévy walk foraging patterns in human huntergatherers, *Proc. Natl. Acad. Sci. USA* **111**, 728 (2014).
- [84] D. W. Sims, A. M. Reynolds, N. E. Humphries, E. J. Southall, V. J. Wearmouth, B. Metcalfe, and R. J. Twitchett, Hierarchical random walks in trace fossils and the origin of optimal search behavior, *Proc. Natl. Acad. Sci. USA* **111**, 11073 (2014).
- [85] G. H. Pyke, Understanding movements of organisms: It's time to abandon the Lévy foraging hypothesis, *Methods Ecol. Evol.* **6**, 1 (2015).
- [86] A. Reynolds, Liberating Lévy walk research from the shackles of optimal foraging, *Phys. Life Rev.* **14**, 59 (2015).
- [87] F. Mery and J. G. Burns, Behavioural plasticity: An interaction between evolution and experience, *Evol. Ecol.* **24**, 571 (2010).
- [88] M. S. Wheatland, P. A. Sturrock, and J. M. McTiernan, The waiting-time distribution of solar flare hard X-ray bursts, *Astrophys. J.* **509**, 448 (1998).
- [89] G. Boffetta, V. Carbone, P. Giuliani, P. Veltri, and A. Vulpiani, Power laws in solar flares: Self-organized criticality or turbulence? *Phys. Rev. Lett.* **83**, 4662 (1999).
- [90] J. M. Beggs and D. Plenz, Neuronal avalanches in neocortical circuits, *J. Neurosci.* **23**, 11167 (2003).
- [91] M. Newman, Power laws, Pareto distributions and Zipf's law, *Contemp. Phys.* **46**, 323 (2005).
- [92] J.-P. Bouchaud and A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, *Phys. Rep.* **195**, 127 (1990).
- [93] D. ben Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, UK, 2000).
- [94] Y. Tu and G. Grinstein, How white noise generates power-law switching in bacterial flagellar motors, *Phys. Rev. Lett.* **94**, 208101 (2005).
- [95] T. S. Biró and A. Jakovác, Power-law tails from multiplicative noise, *Phys. Rev. Lett.* **94**, 132302 (2005).
- [96] I. Lubashevsky, R. Friedrich, and A. Heuer, Realization of Lévy walks as Markovian stochastic processes, *Phys. Rev. E* **79**, 011110 (2009).
- [97] A. M. Reynolds and N. T. Ouellette, Swarm dynamics may give rise to Lévy flights, *Sci. Rep.* **6**, 30515 (2016).
- [98] J. Touboul and A. Destexhe, Power-law statistics and universal scaling in the absence of criticality, *Phys. Rev. E* **95**, 012413 (2017).

- [99] V. Priesemann and O. Shriki, Can a time varying external drive give rise to apparent criticality in neural systems? *PLoS Comput. Biol.* **14**, e1006081 (2018).
- [100] M. Morrell, I. Nemenman, and A. J. Sederberg, Neural criticality from effective latent variable, *eLife* **12**, RP89337 (2023).
- [101] T. Mora and W. Bialek, Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268 (2011).
- [102] L. Cocchi, L. L. Gollo, A. Zalesky, and M. Breakspear, Criticality in the brain: A synthesis of neurobiology, models and cognition, *Prog. Neurobiol.* **158**, 132 (2017).
- [103] J. O’Byrne and K. Jerbi, How critical is brain criticality? *Trends Neurosci.* **45**, 820 (2022).
- [104] D. J. Schwab, I. Nemenman, and P. Mehta, Zipf’s Law and criticality in multivariate data without fine-tuning, *Phys. Rev. Lett.* **113**, 068102 (2014).
- [105] Interestingly, fast-fluctuating hopping rates and scale-invariance arguments have been used to explain heavy-tailed distributions of uncorrelated resting times in mice [128].
- [106] D. G. Hernández, C. Rivera, J. Cande, B. Zhou, D. L. Stern, and G. J. Berman, A framework for studying behavioral evolution by reconstructing ancestral repertoires, *eLife* **10**, e61806 (2021).
- [107] S. Petrovskii, A. Mashanova, and V. A. A. Jansen, Variation in individual walking behavior creates the impression of a Lévy flight, *Proc. Natl. Acad. Sci. USA* **108**, 8704 (2011).
- [108] K. R. Ghusinga, J. J. Dennehy, and A. Singh, First-passage time approach to controlling noise in the timing of intracellular events, *Proc. Natl. Acad. Sci. USA* **114**, 693 (2017).
- [109] X. M. de Wit, A. van Kan, and A. Alexakis, Bistability of the large-scale dynamics in quasi-two-dimensional turbulence, *J. Fluid Mech.* **939**, R2 (2022).
- [110] https://github.com/AntonioCCosta/fluctuating_potential
- [111] A. C. Costa (2024), Fluctuating Landscapes and Heavy Tails in Animal Behavior [Data set], [10.5281/zenodo.10650313](https://zenodo.org/record/10650313).
- [112] J. E. Sulston and S. Brenner, The dna of caenorhabditis elegans, *Genetics* **77**, 95 (1974).
- [113] E. M. Bollt and N. Santitissadeekorn, *Applied and Computational Measurable Dynamics* (SIAM, Philadelphia, 2013).
- [114] A. Bittracher, P. Koltai, S. Klus, R. Banisch, M. Dellnitz, and C. Schütte, Transition manifolds of complex metastable systems, *J. Nonlinear Sci.* **28**, 471 (2018).
- [115] E. Yemini, T. Jucikas, L. J. Grundy, A. E. Brown, and W. R. Schafer, A database of Caenorhabditis elegans behavioral phenotypes, *Nat. Methods* **10**, 877 (2013).
- [116] https://github.com/AntonioCCosta/fluctuating_potential
- [117] L. McInnes, J. Healy, and J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, [arXiv:1802.03426](https://arxiv.org/abs/1802.03426).
- [118] <https://github.com/lmcinnes/umap>
- [119] V. A. Epanechnikov, Non-parametric estimation of a multivariate probability density, *Theory Probab. Appl.* **14**, 153 (1969).
- [120] W. K. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Non-parametric and Semiparametric Models* (Springer, Heidelberg, 2006).
- [121] W. Horsthemke and R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology* (Springer, Heidelberg, 2006).
- [122] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
- [123] G. Pavliotis, *Stochastic Processes and Applications Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, Berlin, 2014).
- [124] S. H. Choi, V. D. Rao, T. Gernat, A. R. Hamilton, G. E. Robinson, and N. Goldenfeld, Individual variations lead to universal and cross-species patterns of social behavior, *Proc. Natl. Acad. Sci. USA* **117**, 31754 (2020).
- [125] We note that for the general dynamics of Eqs. (2) and (3), ω_0 may have an s dependency. However, without loss of generality, we consider that ω_0 and $\Delta U(s)$ can be redefined to move the s dependency to the exponential as a subdominant contribution.
- [126] N. G. van Kampen, Itô versus stratonovich, *J. Stat. Phys.* **24**, 175 (1981).
- [127] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, *Proc. Natl. Acad. Sci. USA* **107**, 11865 (2010).
- [128] A. Proekt, J. R. Banavar, A. Maritan, and D. W. Pfaff, Scale invariance in the dynamics of spontaneous behavior, *Proc. Natl. Acad. Sci. USA* **109**, 10564 (2012).

Correction: Support information in the Acknowledgment section was incomplete and has been fixed.