Switching between phenotypes and population extinction

Ingo Lohmar* and Baruch Meerson†
Racah Institute of Physics, the Hebrew University of Jerusalem, Jerusalem 91904, Israel
(Received 26 July 2011; published 3 November 2011)

Many types of bacteria can survive under stress by switching stochastically between two different phenotypes: the "normals" who multiply fast, but are vulnerable to stress, and the "persisters" who hardly multiply, but are resilient to stress. Previous theoretical studies of such bacterial populations have focused on the *fitness*: the asymptotic rate of unbounded growth of the population. Yet for an isolated population of established (and not very large) size, a more relevant measure may be the population *extinction risk* due to the interplay of adverse extrinsic variations and intrinsic noise of birth, death and switching processes. Applying a WKB approximation to the pertinent master equation of such a two-population system, we quantify the extinction risk, and find the most likely path to extinction under both favorable and adverse conditions. Analytical results are obtained both in the biologically relevant regime when the switching is rare compared with the birth and death processes, and in the opposite regime of frequent switching. We show that rare switches are most beneficial in reducing the extinction risk.

DOI: 10.1103/PhysRevE.84.051901 PACS number(s): 87.18.Tt, 02.50.Ga, 05.40.Ca, 87.23.Kg

I. INTRODUCTION

Understanding and quantifying the persistence of bacterial populations is of major importance for the efficient treatment of diseases. While bacterial persistence was uncovered more than 65 years ago [1], conclusive evidence for the underlying mechanism was only obtained during the last decade from laboratory experiments at the single-cell level. It has been established that an isogenetic population under identical conditions can still exhibit two different phenotypes. They are clearly distinguished by different rates of cell division: "normals" multiply fast and "persisters" do it much slower. For the same reason, however, normals are much more susceptible to antibiotic treatment, while persisters are highly resilient to the antibiotic. An individual bacterium can switch stochastically (at a certain rate, often without sensing its environment) between the two phenotypes [2] (type-II persistence).

Systems of two interacting subpopulations, such as normals and persisters, have been studied in different contexts in theoretical biology [3–6]. Deterministic models of exponential (unbounded) growth were mostly employed, and analysis focused on the *fitness*—the time-averaged net growth rate—of the total population, see, e.g., Refs. [7–11]. In favorable conditions, when normal bacteria have a high net growth rate, frequently switching to persisters is merely a burden, as it decreases the average net growth. If the environment changes (deterministically or randomly) between different states, including some that represent adverse conditions for the normals, e.g., in the presence of an antibiotic, the same frequent switching can become beneficial. In this case, the persisters uphold a base population size during such a stress phase, while normals are heavily decimated. By properly tuning the switching rates between different phenotypic states, one can optimize the fitness of the total population [8]. For two phenotypes and two environments, the average time spent as a certain phenotype should be equal to the average duration of the environment in which this phenotype is the fittest one. In more complicated models (including phenotype-specific response and recovery times upon a change of the environment), one still finds that comparing two (genetic) species, the one with switching rates better tuned (in the above sense) outperforms the other fitnesswise [10].

These are important insights into the role that persisters play in a growing population. However, the underlying assumption of exponential growth is tailored to the description of competition among different genotypes trying to establish themselves by outgrowing others. Here, fitness is instrumental to survive in the competition, and a good indicator of a specific genotypes' prospects. While such an unbounded growth can be realized *in vitro*, the necessary resources and space *in vivo* are limited. To account for this fact, one should introduce models with bounded growth [12]. In a deterministic (mean-field) description, the population will then typically exhibit a stable fixed point corresponding to an established population. In addition, there will be a fixed point describing an extinct population. In reality, population dynamics is a stochastic process: an established population is subject to noise coming from the random character of births and deaths. A rare chain of events, where deaths dominate over births, eventually drives an isolated established population into the absorbing extinction state. Thus for an isolated established population, the ultimate goal is survival in the face of intrinsic, and also, possibly, environmental noise. We suggest, therefore, a paradigm shift in the analysis of bacterial phenotype switching by focusing on the population extinction risk.

With this motivation, we consider a simple two-population system of normals and persisters, possibly in a time-varying environment mimicking a phase of catastrophic conditions for the population. In a constant environment, a proper measure of the extinction risk is the mean time to extinction (MTE) of the population, see, e.g., Ref. [13]. We show that a higher fraction of persisters *exponentially* increases the MTE even in this setting. With a transient catastrophic phase, a more informative measure of extinction risk is the extinction probability increase (EPI) because of the catastrophe [14]. Here, a higher fraction of persisters exponentially reduces

*lohmar@phys.huji.ac.il
†meerson@cc.huji.ac.il

the EPI. Therefore, when viewed from the perspective of population extinction risk, the presence of persisters is always beneficial, providing an "insurance policy" against extinction in small communities. This should be compared with persisters being a mere burden, unless in adverse conditions, when viewed from the perspective of fitness.

The remainder of the paper is organized as follows. In Sec. II, we set up a simple model that describes the interacting populations of normals and persisters. We also introduce, in the same section, the pertinent master equation and employ a WKB approximation, which reduces the master equation to an effective Hamiltonian mechanics. We formulate the mechanical problem that needs to be solved and describe a numerical iteration method for dealing with this problem. Section III presents a perturbation theory, based on time-scale separation, first for favorable conditions, then including a catastrophic phase. There we obtain approximate analytic results for the MTE or the EPI, respectively, and for the most probable path to extinction, and compare them with our numerical solutions. In Sec. IV, we contrast the biologically relevant regime of rare switching with the regime of frequently-switching bacteria. We discuss the main findings in Sec. V.

II. MODEL AND METHODOLOGY

We consider a well mixed two-population system the dynamics of which is described by a continuous-time Markov process. The number of "normals" is denoted by n and that of "persisters" by m. Normals die at a rate that we set to unity throughout, and they multiply at a rate $B(1-n/N)$ per individual. In a stochastic model, this corresponds to a finite state space, with a maximum number $n = N$ of normal individuals. N can be thought of as a number of sites each of which can carry at most one individual or as food resources necessary to produce offspring. This dynamics coincides with that of infected individuals in the SIS model, with a fixed total population size N, unit recovery rate of infected, and an infection rate B/N between infected and susceptible individuals [15].

We now introduce a persister population whose individuals do not multiply or die at all. The populations are coupled by normal individuals switching to persisters at a rate α, and persisters switching to normals at a rate β. The ratio of these switching rates is denoted $\Gamma = \alpha/\beta$. In a mean-field description, the average numbers of individuals are governed by rate equations

$$\begin{aligned} \dot{n} &= Bn(1-n/N) - n - \alpha n + \beta m, \\ \dot{m} &= \alpha n - \beta m. \end{aligned} \tag{1}$$

The rate equations have a trivial fixed point (FP) F_0 at $n = m = 0$, which describes population extinction, and a nontrivial FP $F_{\rm M}$ at $n_{\rm M} = N(1-1/B)$, $m_{\rm M} = \Gamma n_{\rm M}$. A viable population therefore needs $B > 1$, when $F_{\rm M}$ is stable, while F_0 is a saddle point. At the stable FP $F_{\rm M}$, the ratio between the population sizes of persisters and normals is Γ.

A. Noise and metastability

Even for large population size, intrinsic noise is crucial, as it will ultimately drive the system, residing in the vicinity of the deterministically stable FP $F_{\rm M}$, toward extinction. The stochastic system is described by the master equation for the dynamics of the probability distribution of population sizes, $\mathcal{P}_{n,m}(t)$,

$$\begin{aligned} \frac{{\rm d}\mathcal{P}_{n,m}}{{\rm d}t} = \hat{H}\mathcal{P}_{n,m} &= B(n-1)\left(1-\frac{n-1}{N}\right)\mathcal{P}_{n-1,m} \\ &- Bn\left(1-\frac{n}{N}\right)\mathcal{P}_{n,m} + (n+1)\mathcal{P}_{n+1,m} - n\mathcal{P}_{n,m} \\ &+ \alpha(n+1)\mathcal{P}_{n+1,m-1} - \alpha n\mathcal{P}_{n,m} \\ &+ \beta(m+1)\mathcal{P}_{n-1,m+1} - \beta m(1-\delta_{n,N})\mathcal{P}_{n,m}. \end{aligned} \tag{2}$$

Here, the Kronecker delta $\delta_{n,N}$ prevents transition to a state with $n = N+1$. Together with the prescription $\mathcal{P}_{n<0,m} = 0 = \mathcal{P}_{n,m<0}$ and $\mathcal{P}_{n>N,m} = 0$, probability is conserved and limited to the stripe $(n,m) \in [0,N] \times [0,\infty)$. The extinction probability $\mathcal{P}_{0,0}(t)$ is described by the equation

$$\frac{{\rm d}\mathcal{P}_{0,0}}{{\rm d}t} = \mathcal{P}_{1,0}. \tag{3}$$

When higher moments are assumed to factorize, the mean-field equations (1) are recovered by summation over Eq. (2).

The stochastic system, as described by Eq. (2), has an absorbing extinction state $n = 0 = m$, corresponding to zero eigenvalue and eigenstate $\delta_{n,0;m,0}$ of the transition matrix $\hat{H}$. All other eigenvalues are negative, hence all other eigenstates of the probability distribution decay, and the population goes extinct. We assume (and verify *a posteriori*) that, in contrast to all other nonzero eigenvalues, the eigenvalue with smallest nonzero absolute value is exponentially small in the system size N. This corresponds to a *metastable* distribution centered around $F_{\rm M}$ [14,16–25]. The shape function of this distribution, normalized to unity, is called the quasistationary distribution (QSD); we denote it by $\pi_{n,m}$. The decay time of the metastable distribution is $\tau \gg 1$. An initial distribution, describing a viable population, first quickly relaxes to the QSD on a time scale $\sim 1/(B-1)$. Then the metastable distribution will "leak" to zero, as described by the equations $\mathcal{P}_{n,m}(t) \simeq \pi_{n,m}\exp(-t/\tau)$ [for $(n,m) \neq (0,0)$] and $\mathcal{P}_{0,0}(t) \simeq 1 - \exp(-t/\tau)$, where τ is expected to be exponentially large in N. Using Eq. (2), the QSD $\pi_{n,m}$ obeys the equation

$$\hat{H}\pi_{n,m} = -\pi_{n,m}/\tau, \tag{4}$$

and with τ exponentially large in N, the right-hand side can be approximated by zero. Having found $\pi_{n,m}$, one obtains τ by using Eq. (3): $\tau = 1/\pi_{1,0}$. One can show (see, e.g., Ref. [19]) that τ is indeed the mean time to extinction (MTE) when starting from the QSD. We remind the reader that time is measured throughout this paper in units of the death rate coefficient of the normal population.

B. WKB approximation

When N is sufficiently large, one can approximately solve Eq. (4) by a Wentzel-Kramers-Brillouin (WKB) eikonal ansatz [16,26–28]:

$$\pi_{n,m} = \exp\left[-NS(x,y)\right], \tag{5}$$

where $x = n/N$ and $y = m/N$ are assumed to be continuous variables. Having found $S(x,y)$ in the leading order in

$1/N$, the MTE can be calculated up to a pre-exponential factor:

$$\tau = 1/\pi_{1,0} \approx \exp[N S(0,0)], \tag{6}$$

such that $S(0,0)$ plays the role of an entropic barrier against extinction.

Plugging Eq. (5) into Eq. (4) and Taylor-expanding S around (x,y) to first order, one obtains, in the leading order of $1/N$, a zero-energy Hamilton-Jacobi equation

$$H(x,y,\partial S/\partial x, \partial S/\partial y) = 0, \tag{7}$$

where

$$H(x,y,p_x,p_y) = Bx(1-x)(e^{p_x}-1) + x(e^{-p_x}-1) + \alpha x(e^{-p_x+p_y}-1) + \beta y(e^{p_x-p_y}-1) \tag{8}$$

is the effective Hamiltonian. The corresponding Hamilton equations,

$$\dot{x} = Bx(1-x)e^{p_x} - xe^{-p_x} - \alpha x e^{-p_x+p_y} + \beta y e^{p_x-p_y}, \tag{9a}$$

$$\dot{y} = \alpha x e^{-p_x+p_y} - \beta y e^{p_x-p_y}, \tag{9b}$$

$$\dot{p}_x = -B(1-2x)(e^{p_x}-1) - (e^{-p_x}-1) - \alpha(e^{-p_x+p_y}-1), \tag{9c}$$

$$\dot{p}_y = -\beta(e^{p_x-p_y}-1), \tag{9d}$$

describe trajectories of the system in the four-dimensional phase space of rescaled population sizes x and y and conjugate momenta p_x and p_y. To determine $S(x,y)$, one can calculate the mechanical action accumulated along the proper activation trajectory, or *instanton*, of Hamilton's equations of motion and ending in (x,y).

As the Hamiltonian H does not explicitly depend on time, $H(x,y,p_x,p_y) = E$ is an integral of motion. In view of Eq. (7), the energy E must be zero. One type of motion with $E = 0$ occurs in the invariant plane $p_x = p_y = 0$ where Eqs. (9a) and (9b) coincide with the (rescaled) rate equations (1). Overall, there are three zero-energy FPs of the Hamiltonian flow: (0,0,0,0), $[1-1/B, \Gamma(1-1/B), 0, 0]$, and $(0, 0, -\ln B, -\ln B)$, all of them four-dimensional saddles. The first two originate from the mean-field FPs, and we will continue referring to them as F_0 and F_M, respectively. The third FP, which we call $F_\emptyset$, is the fluctuational extinction point: it appears in a broad class of stochastic population models exhibiting extinction [17,20,21,29]. Note that all the FPs merge into the origin upon approaching the bifurcation point $B = 1$.

As the established population resides around F_M, the instanton must start at this FP. Now, as we look for $S(0,0)$, we need to choose between the fixed points F_0 and $F_\emptyset$ as the final destination. It has been shown that only $F_\emptyset$ can be reached from the region $x,\ y > 0,\ p_x,\ p_y \neq 0$ [21,24]. The instanton, therefore, must be a heteroclinic trajectory, which starts at the metastable FP F_M at time $-\infty$ and enters the extinction FP $F_\emptyset$ at time $+\infty$. Finding the MTE, see Eq. (6), demands calculating the action $S = S(0,0)$ along this heteroclinic trajectory:

$$S = \int \mathrm{d}t\,(\mathbf{p}\dot{\mathbf{q}} - H) = \int \mathrm{d}t\,(-\dot{\mathbf{p}}\mathbf{q} - H) = \int (p_x\,\mathrm{d}x + p_y\,\mathrm{d}y - H\,\mathrm{d}t), \tag{10}$$

where $\mathbf{q} = (x,y)$ and $\mathbf{p} = (p_x, p_y)$. In a boundary layer of width $\sim 1/N$ around $x = 0$ and $y = 0$ the assumption of large population size $n,\ m \gg 1$ breaks down. However, for a sufficiently large system size N, the contribution of this layer to the MTE is subleading in the parameter $1/N$ [25,30].

C. Iterative numerical solution

The two-degrees-of-freedom Hamiltonian (8) has only one independent integral of motion: the energy. It is thus nonintegrable. Therefore the instanton can in general be only obtained numerically.

In earlier work, "shooting" algorithms were used to integrate numerically Hamilton's equations of motion for this purpose, see, e.g., Refs. [14,20,21]. Below (Sec. III C) we will explain why such an algorithm is not feasible in our case. Instead, we adapted an iterative algorithm introduced, in the context of Hamiltonian field theories, in Refs. [18,31]. Let subscripts "M" and "ø" label the initial and the final FP, respectively. We fix a sufficiently long calculation time $t_{\max}$ to traverse the trajectory; it should not be too long in order to avoid instabilities in the vicinities of the fixed points. The starting iteration numerically integrates Eqs. (9a) and (9b) with the momenta fixed at their target values $\mathbf{p} = \mathbf{p}_\emptyset$, starting from the initial condition $\mathbf{q}(t=0) = \mathbf{q}_M$ and up to time $t_{\max}$. The resulting coordinate curve $\mathbf{q}(t)$ is now used to fix the coordinates in Eqs. (9c) and (9d), leaving a system of equations for the momenta, which is integrated backward in time starting from $\mathbf{p}(t = t_{\max}) = \mathbf{p}_\emptyset$ down to $t = 0$. In each following iteration half step, momenta (coordinates) are fixed by the time-dependent solution obtained in the previous step, and the coordinates (momenta) are integrated forward (backward) in time, starting from the values at the initial (final) FP and up (down) to $t = t_{\max}$ $(t = 0)$. We found that this scheme rapidly converges to the desired instanton.

To compute the action, we use the expressions in the first line of Eq. (10). The difference between these two versions is an easy measure of the numerical accuracy that has been reached. This algorithm makes it possible to obtain, with little effort, the most likely path to extinction and the MTE for a broad class of population dynamics models when the target FP has a different momentum than the initial FP (as it happens here).

III. INSTANTON TRAJECTORIES

A. Close to the bifurcation

To simplify the algebra, we will restrict ourselves to the regime close to the bifurcation point $B = 1$ where all FPs merge, and define the distance to bifurcation $\delta = B - 1 \ll 1$. As can be checked *a posteriori*, $x,\ y/\Gamma,\ |p_x|,\ |p_y| \sim \delta$ or smaller. Therefore exponentials in the Hamiltonian (8) can be Taylor-expanded. In addition, we assume that the switching

from the normals to persisters and back is rare: $\alpha, \beta \ll \delta \ll 1$. Under these conditions, the Hamiltonian (8) becomes

$$H(x,y,p_x,p_y) \simeq x p_x(p_x - x + \delta) - (\alpha x - \beta y)(p_x - p_y). \tag{11}$$

Here, we neglected terms $\sim \delta^4$, and the term $(\alpha x + \beta y)(p_x - p_y)^2/2 \sim \alpha\delta^3$. This is consistent if $\alpha\delta^2 \gg \delta^4$, that is, $\delta \ll \sqrt{\alpha}$. The Hamilton equations read

$$\dot{x} = x(2p_x - x + \delta) - (\alpha x - \beta y), \tag{12a}$$
$$\dot{y} = \alpha x - \beta y, \tag{12b}$$
$$\dot{p}_x = -p_x(p_x - 2x + \delta) + \alpha(p_x - p_y), \tag{12c}$$
$$\dot{p}_y = -\beta(p_x - p_y), \tag{12d}$$

and the zero-energy FPs are (0,0,0,0) (trivial FP, F_0), (0,0, $-\delta$, $-\delta$) (extinction FP, $F_\emptyset$), and (δ,$\Gamma\delta$,0,0) (metastable FP, F_M).

It is helpful to rescale all quantities by putting $x = \delta X$, $y = \delta Y$, $p_x = \delta P_X$, $p_y = \delta P_Y$, and $t = T/\delta$. The equations of motion become

$$\frac{dX}{dT} = X(2P_X - X + 1) - \varepsilon(\Gamma X - Y), \tag{13a}$$
$$\frac{dY}{dT} = \varepsilon(\Gamma X - Y), \tag{13b}$$
$$\frac{dP_X}{dT} = -P_X(P_X - 2X + 1) + \varepsilon\Gamma(P_X - P_Y), \tag{13c}$$
$$\frac{dP_Y}{dT} = -\varepsilon(P_X - P_Y), \tag{13d}$$

where $\varepsilon = \beta/\delta$. These equations are still canonical with Hamiltonian

$$h = H/\delta^3 = X P_X(P_X - X + 1) - \varepsilon(\Gamma X - Y)(P_X - P_Y). \tag{14}$$

The action becomes $S = \delta^2 s$, where

$$s = \int (P_X\, dX + P_Y\, dY - h\, dT). \tag{15}$$

The rare-switching limit corresponds to $\varepsilon \ll 1$, and we will treat it perturbatively in the following.

B. Solution in a constant favorable environment

The leading-order behavior of X and P_X, the *fast* degrees of freedom, takes place on the unit time scale $T \sim 1$. The dynamics of Y and P_Y, the *slow* degrees of freedom, however happens on the long time scale $T \sim 1/\varepsilon \gg 1$. We formally introduce a separate slow time variable $T' = \varepsilon T$ to account for this separation of time scales, and consider perturbative solutions of the form

$$\begin{aligned} X &= X_0(T) + \varepsilon X_1(T,T') + \dots, \\ P_X &= P_{X0}(T) + \varepsilon P_{X1}(T,T') + \dots, \\ Y &= Y_0(T') + \varepsilon Y_1(T') + \dots, \\ P_Y &= P_{Y0}(T') + \varepsilon P_{Y1}(T') + \dots. \end{aligned} \tag{16}$$

Inserting into the Hamilton equations (13) yields a system of partial differential equations in each order of ε. Note that, in contrast to previous work [20,21], here the dynamics of fast variables (normals) drives the slow variables (persisters).

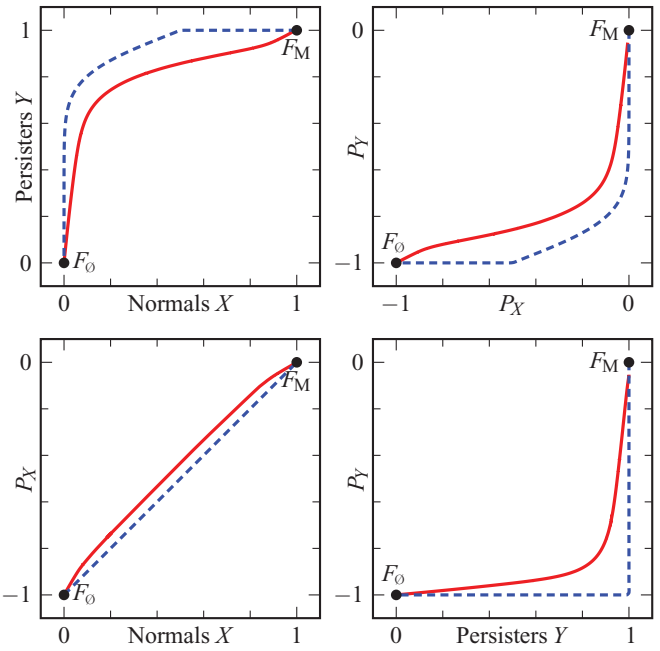

FIG. 1. (Color online) Instanton (constant environment, close to bifurcation) for $\Gamma = 1$ and $\varepsilon = 0.1$ in several projections. Theory prediction (dashed blue) and numerical solution (solid red).

In the leading order $\sim \varepsilon^0$, only two equations remain, $dX_0/dT = X_0(2P_{X0} - X_0 + 1)$ and $dP_{X0}/dT = -P_{X0}(P_{X0} - 2X_0 + 1)$. This amounts to the one-dimensional system of [14,29] close to the bifurcation. The solution must satisfy the energy constraint $h_{X0} = X_0 P_{X0}(P_{X0} - X_0 + 1) = 0$, hence $P_{X0} = X_0 - 1$: the projection of the instanton to the X-P_X plane is a straight line between F_M and $F_\emptyset$ (cf. Fig. 1), and this part contributes an action $s_{X0} = 1/2$ [14,20,21]. The solutions for X_0 and P_{X0} are

$$X_0(T) = \frac{1}{1 + e^T}, \quad P_{X0}(T) = \frac{-1}{e^{-T} + 1}, \tag{17}$$

where we have arbitrarily fixed the position of the instanton along the time axis.

The slow persister variables appear in the order $\sim \varepsilon^1$,

$$\begin{aligned} \frac{dY_0}{dT'} + Y_0(T') &= \Gamma X_0(T), \\ \frac{dP_{Y0}}{dT'} - P_{Y0}(T') &= -P_{X0}(T). \end{aligned} \tag{18}$$

On the slow time scale of the left-hand sides, the driving terms $X_0(T)$ and $P_{X0}(T)$ change with time only in the narrow region $|T'| \sim \varepsilon \ll 1$; for earlier and later times, they are almost constant. Therefore, on the slow time scale, they can be described as step functions $X_0 = \theta(-T')$ and $P_{X0} = -\theta(T')$. We thus solve $dY_0/dT' + Y_0(T') = \Gamma\theta(-T')$ by matching solutions [with $Y_0(-\infty) = \Gamma$, $Y_0(+\infty) = 0$] at $T' = 0$,

$$Y_0(T') = \begin{cases} \Gamma & \text{for } T' \leqslant 0, \\ \Gamma e^{-T'} & \text{for } T' \geqslant 0. \end{cases} \tag{19}$$

Similarly, we have $dP_{Y0}/dT' - P_{Y0}(T') = \theta(T')$ [with $P_{Y0}(-\infty) = 0$, $P_{Y0}(+\infty) = -1$], such that

$$P_{Y0}(T') = \begin{cases} -e^{T'} & \text{for } T' \leqslant 0, \\ -1 & \text{for } T' \geqslant 0. \end{cases} \tag{20}$$

The phase trajectory projection to the Y-P_Y plane forms a rectangle and contributes an area $s_{Y0} = \Gamma$ to the action. To resolve the small region $|T'| \lesssim \varepsilon$, one would need to include subleading corrections, which would smoothen the discontinuous derivatives of Y_0 and P_{Y0} at $T' = 0$, round off the trajectory, and decrease the action by small terms $\sim \varepsilon$.

The total action in the leading order $\sim \varepsilon^0$ reads

$$s_0 = \tfrac{1}{2} + \Gamma. \tag{21}$$

The MTE of the population becomes, up to a pre-exponent,

$$\tau \simeq \exp(N\delta^2 s_0) = \exp\left[N\delta^2\left(\frac{1}{2} + \Gamma\right)\right]. \tag{22}$$

In comparison, without persisters, the MTE is $\simeq \exp(N\delta^2 s_{X0}) = \exp(N\delta^2/2)$, so the persisters cause an exponential increase of the MTE of the population. A part of the exponential increase comes simply from an increased metastable population size: persisters do not compete with normals, so there is no "cost" of increasing their population (via Γ), only a benefit against extinction. Therefore let us compare the MTE (22) with the MTE τ^{1d} of a single-population system of normals, compensated by $N \to N(1+\Gamma)$. Both systems then have the same carrying capacity $K = N\delta(1+\Gamma)$. The ratio of the MTEs is

$$\frac{\tau}{\tau^{1d}} = \exp\left[\frac{K\delta\Gamma}{2(1+\Gamma)}\right], \tag{23}$$

still exponentially large at $K\delta \gg 1$ and not too small Γ. Notable is the effect of increasing the persister fraction $\Gamma/(1+\Gamma)$, which saturates at large Γ. Equation (23) does not suggest any optimal value of Γ but the largest possible one; we will discuss the relation to other results and the biological context in Sec. V.

Interestingly, persisters contribute an action that does not depend on the absolute switching rates α and β, see Eq. (21). It may be surprising that an arbitrarily small but finite perturbation $\varepsilon > 0$ yields an exponential change in the MTE with respect to $\varepsilon = 0$. This is yet another instance of extinction rate fragility [22]. As in other "fragile" population systems, the explanation to this counter-intuitive effect comes from a time-resolved picture [23]. The effective extinction rate is time dependent. At relatively small times $1 \ll T \ll 1/\varepsilon$, the extinction rate is the same as if the persisters were absent ($\varepsilon = 0$). At longer times $T \gtrsim 1/\varepsilon$, the extinction rate crosses over to its asymptotic value which determines the MTE (22) [23].

In deriving Eq. (22), we assumed closeness to the bifurcation and rare switching, i.e., $\alpha, \beta \ll \delta \ll 1$, or equivalently ε, $\varepsilon\Gamma$, and $\delta \ll 1$; in particular, implying the upper bound $\Gamma \ll 1/\varepsilon$. To obtain the approximate Hamiltonian (11), we also had to demand $\alpha \gg \delta^2$ ($\varepsilon\Gamma \gg \delta$); with hindsight this can be lifted: solving the (effectively one-dimensional) fast subsystem only employs $\delta \ll 1$, while the ansatz (16) only relies on time-scale separation $\varepsilon \ll 1$. As the small parameters δ and ε describe unrelated mechanisms, the analytical results do not depend (to the given order) on $\varepsilon\Gamma \gg \delta$. The WKB approximation is valid, and the resulting MTE $\tau \gg 1$ is exponentially large, if $N\delta^2(1/2+\Gamma) \gg 1$. For that, a minimum system size $N \gg \delta^{-2}$ is sufficient, when $N^{-1/2} \ll \delta \ll 1$ (QSD width much smaller than the distance between initial and target FPs).

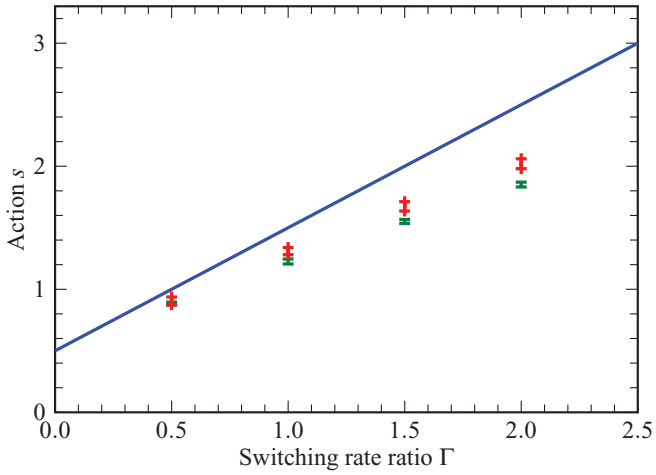

FIG. 2. (Color online) Action s of Eq. (15) vs the ratio of switching rates Γ, analytical (21) (solid blue line) and numerical result (green marks $\varepsilon = 0.2$, red pluses $\varepsilon = 0.1$). The error bars were obtained by using the original action expression and its integrated-by-parts counterpart [see Eq. (10)].

Figure 1 compares the instanton found analytically with the numerical solution (see Sec. II C) of Eqs. (13) for a moderately small $\varepsilon = 0.1$. Agreement is reasonably good, and we checked that it improves, in all projection planes, with decreasing ε. Figure 2 shows that the numerically obtained action approaches the theoretical value (21) as $\varepsilon \to 0$. The deviation also decreases as Γ goes down, as expected.

C. Effect of a catastrophe

What is the effect of a "catastrophe," i.e., temporary adverse conditions, on the population extinction risk? For a single population, this question was addressed in Ref. [14]. Here, we find that the presence of a persister subpopulation dramatically reduces the extinction probability increase (EPI) caused by the same type of catastrophe.

As in Ref. [14], we will model a catastrophe by setting $B = 0$ during a certain period of time t_c. This may mimic the effect of a drug that inhibits cell multiplication. The system history then differs from the one described in Sec. II A. For early times, after relaxation of the system to the QSD, the extinction probability still increases with time nearly linearly as $\mathcal{P}_{0,0}(t) \simeq 1 - \exp(-t/\tau) \simeq t/\tau$, where τ is the MTE of the system without a catastrophe. At a time $t_0 \ll \tau$, when $\mathcal{P}_{0,0} = \mathcal{P}^{\rm pre}_{0,0}$, the catastrophe starts, acting for a duration $t_c \ll \tau$. Compared with τ, this is a short transient, which may however considerably increase the extinction probability to the value $\mathcal{P}^{\rm post}_{0,0}$. Afterwards, the system is again described by the (downscaled) QSD and continues to decay, while the extinction probability increases as $\mathcal{P}_{0,0}(t) \simeq 1 - (1 - \mathcal{P}^{\rm post}_{0,0})\exp[(t_0 + t_c - t)/\tau]$. In this setting, the MTE is too crude a measure of the effect of the catastrophe: it is dominated by realizations surviving the catastrophe, resulting in nearly the unperturbed MTE τ. Instead, we measure the influence of the catastrophe by the EPI $\Delta\mathcal{P}_{0,0} = \mathcal{P}^{\rm post}_{0,0} - \mathcal{P}^{\rm pre}_{0,0}$. Up to a pre-exponential factor it is given by

$$\Delta\mathcal{P}_{0,0} \simeq e^{-NS_c}, \tag{24}$$

where S_c is the mechanical action accumulated along the instanton [14], see Eq. (10). While it describes a very different quantity, one gets, in the leading order, $\Delta\mathcal{P}_{0,0}$ from the action exactly as one gets $1/\tau$ in a constant environment, cf. Eq. (6). For Eq. (24) to be valid, in addition to $NS_c \gg 1$ one has to demand that the change of the exponent with respect to the constant-environment case is large, $N(S - S_c) \gg 1$ [14].

The instanton itself is obtained analogously to the case of time-independent transition rates described in Sec. II B. The Hamiltonian now explicitly depends on time: before and after the catastrophe, the system is still described by the Hamiltonian (8). During the catastrophe, the effective Hamiltonian becomes

$$H_c = x(e^{-p_x} - 1) + \alpha x(e^{-p_x+p_y} - 1) + \beta y(e^{p_x-p_y} - 1). \tag{25}$$

The instanton trajectory now consists of three connected segments: the precatastrophe segment starts at the metastable FP F_M and is determined by the Hamiltonian (8), the catastrophe segment is described by Eq. (25), and the postcatastrophe segment leads to the extinction FP $F_\emptyset$, again governed by Eq. (8). We assume that, after the catastrophe ends, there is still a relatively large population left (with exponentially long MTE). Neither H nor H_c depend on time explicitly, therefore on each segment, energy is conserved: before and after the catastrophe, $H = E = 0$, and during the catastrophe $H_c = E_c \neq 0$. Furthermore, the phase space points matching the segments are fixed by the catastrophe duration t_c. In turn, this fixes the energy E_c.

Again, we consider the system close to the bifurcation, $\delta \ll 1$, and assume rare switching, $\alpha, \beta \ll \delta \ll 1$, such that before and after the catastrophe we have the Hamiltonian (11). We expect (and check *a posteriori*) that $x, y/\Gamma, |p_x|, |p_y| \sim \delta$ or smaller. This leads to

$$H_c \simeq -xp_x + \frac{xp_x^2}{2} - (\alpha x - \beta y)(p_x - p_y), \tag{26}$$

where we have kept the same orders as for Eq. (11).

Rescaling all quantities by δ as in Sec. III A, the Hamiltonian during the catastrophe becomes

$$h_c = \frac{H_c}{\delta^3} = -\frac{XP_X}{\delta} + \frac{XP_X^2}{2} - \varepsilon(\Gamma X - Y)(P_X - P_Y), \tag{27}$$

with the equations of motion

$$\frac{dX}{dT} = -\frac{X}{\delta} + XP_X - \varepsilon(\Gamma X - Y), \tag{28a}$$

$$\frac{dY}{dT} = \varepsilon(\Gamma X - Y), \tag{28b}$$

$$\frac{dP_X}{dT} = \frac{P_X}{\delta} - \frac{P_X^2}{2} + \varepsilon\Gamma(P_X - P_Y), \tag{28c}$$

$$\frac{dP_Y}{dT} = -\varepsilon(P_X - P_Y). \tag{28d}$$

The rescaled duration of the catastrophe is denoted $T_c = \delta t_c$. The leading terms in dX/dT and dP_X/dT are $\sim 1/\delta \gg 1$: during the catastrophe the population size decays exponentially on the fast time scale.

To get some insight into the impact of the catastrophe, let us consider a numerical solution. To this end, we use the method described in Sec. II C, where the equations of motion

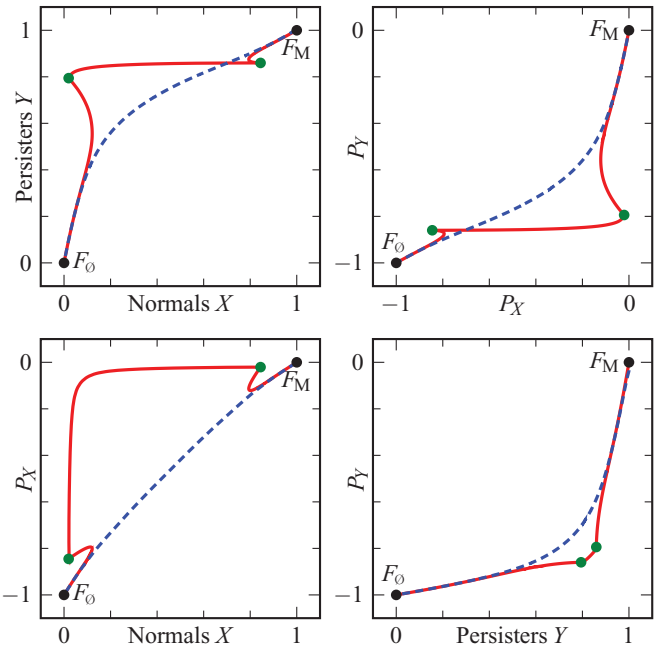

FIG. 3. (Color online) Numerically found instanton for $\Gamma = 1$, $\varepsilon = 0.2$, and $\delta = 0.1$, without a catastrophe (dashed blue) and with a catastrophe of duration $T_c = 0.5$ (solid red). Green dots mark both the start and the end of the catastrophe.

now change from Eq. (13) to Eq. (28) at some time (and back after a duration T_c). The result is insensitive to this starting time, provided it is sufficiently far from $t = 0$ and $t = t_{max}$. Figure 3 shows several projections of an instanton with and without the catastrophe phase, for otherwise identical parameters. In the top panels, due to time-scale separation, the catastrophe segment is nearly horizontal—X and P_X rapidly decay, persisters are (indirectly) affected much later. The bottom panels show that a subpopulation size and its conjugate momentum do not change simultaneously. For persisters, first the momentum builds up, then the population size drops, as in a constant environment (see Fig. 1). For normals, on the other hand, the situation has changed; the population size now decays earlier than the momentum, this will be explained in Sec. III D. The sudden onset and end of the catastrophe is reflected by nonsmoothness of the instanton (except for the Y-P_Y projection). "Wiggles" due to nonmonotonic X and P_X immediately precede or follow the catastrophe segment (we confirmed that these are not numerical artifacts). One can see that, after an initial decay of the normal subpopulation size, it briefly recovers, only to be hit all the harder by the catastrophe. Afterwards, there is a short recovery period caused by influx from persisters (cf. the Y-P_Y projection).

The two-population system with a catastrophe shows a fundamental difference from the single-population case: the instanton is not only changed during the catastrophe phase, but the whole trajectory including pre- and postcatastrophe segments is affected. This can be understood via the following counting argument.

Imagine we try to match, in a d-population system with piecewise constant Hamiltonian, the three segments of the instanton. The $2d$-saddle F_M affords a d-dimensional unstable manifold of possible end points of the precatastrophe segment.

A useful parametrization of this point (where the catastrophe segment begins) consists in $d-1$ "angles" describing different trajectories, and a timelike parameter along the trajectories. By matching the catastrophe segment of a given duration, the phase space point at its end is then fixed as well. At the other end, $F_\emptyset$ affords a d-dimensional stable manifold of possible starting points for the postcatastrophe segment, which can be parametrized as above. We thus have $d+d$ parameters at our disposal describing the possible points at the end of the catastrophe segment and the start of the postcatastrophe segment. Since we have to match them in $2d$ phase space coordinates, this picture does not contradict a unique instanton (although there still may be more than one solution).

For the single-population case $d=1$, the phase trajectories leaving and entering the fixed points are unique, and thus cannot be affected by the catastrophe part in between. In the generic case $d \geqslant 2$, however, the pre/postcatastrophe segments may differ from the no-catastrophe instanton. In the concrete model studied here, these segments have to differ simply because of time-scale separation. During the catastrophe, normals are rapidly decimated, whereas the persister dynamics follows much more slowly. In the X-Y-projection, the catastrophe segment is thus less steep than the slope between any two points on the no-catastrophe instanton. It is therefore impossible to simply splice the catastrophe segment into the latter.

This explains why "shooting" algorithms are impractical for finding the catastrophe-related instanton numerically in a multipopulation system. For a single population with catastrophe [14], such an algorithm can start with a small displacement from the metastable FP along the no-catastrophe instanton, testing different starting points of the catastrophe segment—this works as the precatastrophe segment is unchanged. Likewise, one can parametrize the zero-energy trajectories leaving the initial FP in the two-population system without catastrophe (see Sec. III B) by a shooting angle. Adding a catastrophe provides an additional freedom (in the form of the starting point), and the method is no longer practical.

At the same time, Fig. 3 shows that the instantons without and with catastrophe practically coincide (in all projections) for an extended part next to both FPs, before eventually departing from each other. This means that the system is extremely sensitive to minute variations in the angle at which the trajectory leaves (enters) the initial (final) FP, which only become visible closer to the catastrophe segment. We confirmed this behavior in tests of the aforementioned shooting algorithm (without catastrophe), which, for this reason, already proves to be rather tedious.

D. Analytic theory with catastrophe

We look for an analytic solution analogously to Sec. III B. Time-scale separation is still effective: X and P_X show fast dynamics on the time scale $T \sim 1$, or even $T \sim \delta$, see Eq. (28). They drive the slow Y and P_Y, which change on a time scale $T' = \varepsilon T$. We denote the catastrophe duration on this scale by $T'_c = \varepsilon T_c$.

The leading-order equations $\sim \varepsilon^0$ reduce to the normals-only system again, and $h_{X,c} = -XP_X/\delta + XP_X^2/2$ governs the dynamics during the catastrophe. Since $X, P_X \sim 1 \ll 1/\delta$, we neglect the second term, and arrive at the simple catastrophe Hamiltonian $h_{X,c} \simeq -XP_X/\delta$ used in the single-population model [14]. The solution is an exponential decay (growth) of X (P_X) at a rate $1/\delta$ and for a duration T_c. Let $X^+ > X^-$ and $0 > P_X^+ > P_X^-$ denote coordinates and momenta at the start and the end of the catastrophe, respectively. Then $X^- = X^+ \exp(-T_c/\delta)$ and $P_X^- = P_X^+ \exp(+T_c/\delta)$. The solution for X and P_X before and after the catastrophe is the same (up to a time shift) as in the constant environment, Sec. III B. This is no contradiction to the arguments of Sec. III C, since the leading approximation is effectively one-dimensional. Therefore

$$X_0(T) = \begin{cases} (1+e^{T-T_<})^{-1} & \text{for } X_0 \geqslant X^+, \\ (1+e^{T-T_>})^{-1} & \text{for } X_0 \leqslant X^-, \end{cases} \tag{29}$$

and $P_{X0}(T) = X_0(T) - 1$ for both $P_{X0} \geqslant P_X^+ = X^+ - 1$ and $P_{X0} \leqslant P_X^- = X^- - 1$. The quantities $T_<$ and $T_>$ are yet undetermined. From the constraints, we get

$$X^\pm = \frac{1}{1+e^{\mp T_c/\delta}}, \quad P_X^\pm = \frac{-1}{1+e^{\pm T_c/\delta}}, \tag{30}$$

and the conserved (X-part) energy during the catastrophe becomes $h_{X0,c} = \cosh^{-2}[T_c/(2\delta)]/(4\delta)$. Fixing the time such that the catastrophe occurs between $T = \pm T_c/2$, we obtain

$$X_0(T) = \begin{cases} (1+e^{T-T_c(1/\delta-1/2)})^{-1} & \text{for } T \leqslant -\frac{T_c}{2}, \\ \dfrac{\exp(-T/\delta)}{2\cosh[T_c/(2\delta)]} & \text{for } -\frac{T_c}{2} \leqslant T \leqslant +\frac{T_c}{2}, \\ (1+e^{T+T_c(1/\delta-1/2)})^{-1} & \text{for } \frac{T_c}{2} \leqslant T. \end{cases} \tag{31}$$

The momentum is $P_{X0} = X_0 - 1$ before and after the catastrophe, and during it decays as

$$P_{X0}(T) = \frac{-e^{T/\delta}}{2\cosh[T_c/(2\delta)]} = -X_0(-T). \tag{32}$$

The action found for this Hamiltonian and trajectory is $s_{X0,c} = [1+\exp(T_c/\delta)]^{-1}$ [14]. During the catastrophe, the "trajectory contribution" $\int P_{X0}\,\mathrm{d}X_0$ and $\int -h_{X0,c}\,\mathrm{d}T$ cancel each other.

The slow equations of motion (28b) and (28d) are the same as in the favorable environment of Sec. III B, hence the slow leading-order equations (18) (and boundary conditions) are unchanged. Again, we only resolve the slow dynamics here. The driving terms X_0 and P_{X0} are different now, since a part of their movement is replaced by a faster exponential decay (rate $1/\delta \gg 1$) during the catastrophe. Therefore on the slow T' scale, one obtains a step function as an even better approximation than in Sec. III B. The only difference between Eqs. (17) and (31) is that the driving by X_0 (P_{X0}) ends (sets in) at the start (end) of the catastrophe $T = \mp T_c/2$ (instead of $T = 0$), such that $X_0 = \theta(-T'_c/2 - T')$ and $P_{X0} = -\theta(T' - T'_c/2)$.

Since coordinates and momenta remain separate in Eq. (18), the general piecewise solutions for Y_0 and P_{Y0} are unchanged, but now matched at $T' = \mp T'_c/2$:

$$Y_0(T') = \begin{cases} \Gamma & \text{for } T' \leqslant -T'_c/2, \\ \Gamma e^{-T'-T'_c/2} & \text{for } -T'_c/2 \leqslant T', \end{cases} \tag{33}$$

and

$$P_{Y0}(T') = \begin{cases} -e^{T'-T'_c/2} & \text{for } T' \leqslant T'_c/2, \\ -1 & \text{for } T'_c/2 \leqslant T'. \end{cases} \tag{34}$$

The simple geometric picture that the catastrophe merely time-shifts Y_0 and P_{Y0} into opposite directions results in a hyperbola $Y_0 P_{Y0} = -\Gamma \exp(-T_c')$ on the corresponding segment.

Persisters contribute an action

$$s_{Y0,c} = \int P_{Y0}\, dY_0 - \int h_{Y0,c}\, dT, \tag{35}$$

with the switching Hamiltonian $h_{Y,c} = -\varepsilon(\Gamma X - Y)(P_X - P_Y)$. The energy during the catastrophe is evaluated on the slow time scale, such that $X_0 = 0 = P_{X0}$, and

$$h_{Y0,c} = -\varepsilon(\Gamma X_0 - Y_0)(P_{X0} - P_{Y0}) = \varepsilon\Gamma e^{-T_c'}. \tag{36}$$

The contribution to the action $-h_{Y0,c}T_c = -\Gamma T_c' \exp(-T_c')$ again cancels the phase space area under the catastrophe segment, $\int_\Gamma^{\Gamma\exp(-T_c')} P_{Y0}\, dY_0$. Hence the persister action is $s_{Y0,c} = \Gamma\exp(-T_c')$, and the total action becomes

$$s_{0,c} = \frac{1}{1+e^{T_c/\delta}} + \Gamma e^{-T_c'}. \tag{37}$$

Reinstating the original time scale t by using $T' = \varepsilon T = \varepsilon\delta t = \beta t$ we obtain from Eq. (24)

$$\Delta\mathcal{P}_{0,0} \simeq \exp\left[-N\delta^2\left(\frac{1}{1+e^{t_c}} + \Gamma e^{-\beta t_c}\right)\right]. \tag{38}$$

The system without persisters ($\Gamma = 0$) has an EPI $\simeq \exp(-N\delta^2 s_{X0,c}) = \exp[-N\delta^2/(1+e^{t_c})]$. As for favorable conditions, we compare with the EPI $\Delta\mathcal{P}_{0,0}^{1d}$ of such a single-population system of normals, compensated by $N \to N(1+\Gamma)$ to have the same carrying capacity $K = N\delta(1+\Gamma)$:

$$\frac{\Delta\mathcal{P}_{0,0}}{\Delta\mathcal{P}_{0,0}^{1d}} = \exp\left[-\frac{K\delta\Gamma}{1+\Gamma}\left(e^{-\beta t_c} - \frac{1}{1+e^{t_c}}\right)\right]. \tag{39}$$

The system with persisters has exponentially smaller EPI, to which the initial population size K and the persister fraction $\Gamma/(1+\Gamma)$ contribute as to the MTE ratio (23). The parenthesized factor quantifies the fundamental benefit of persisters and generalizes the numerical value $1/2$ in Eq. (23): the effect is most pronounced for catastrophes that are long on the fast scale of normals, but short on the slow persister time scale, $t_c \gg 1 \gg T_c' = \beta t_c$. Then $\Delta\mathcal{P}_{0,0}/\Delta\mathcal{P}_{0,0}^{1d} \simeq \exp[-K\delta\Gamma/(1+\Gamma)]$, i.e., the ratio is squared with respect to the MTE ratio (23) in a constant favorable environment: the benefit of persisters is even more apparent in the face of a catastrophe. Again the result (39) suggests to choose Γ as large as possible, on which we comment in Sec. V.

These results are based on δ, ε, and $\varepsilon\Gamma \ll 1$ (cf. the end of Sec. III B). For a short catastrophe $t_c \sim 1$ or smaller, the WKB result (24) is valid if the reduction $N\delta^2(s_0 - s_{0,c})$ due to the catastrophe is sufficiently large, yielding the condition $N \gg 4\delta^{-2}/t_c$. A long catastrophe $T_c' \sim 1$ (or larger) strongly reduces the action, and the stricter condition is that the remaining action be large enough. Considering $\Gamma \sim 1$ for simplicity, the persister action dominates, leading to $N \gg \exp(T_c')/(\delta^2\Gamma)$.

The theory path to extinction is shown in Fig. 4 and compared with the numerical solution (see Secs. II C and III C). For a short catastrophe $T_c = 0.2$, persisters are mostly unaffected, while the X-P_X projection resembles the one-dimensional system [14]. Already for the moderate $T_c = 1$ (not shown), normals have gone virtually extinct at the end of the catastrophe, and the population survives mainly due to the remaining persisters. With a long catastrophe $T_c = 10$, the action contributed by persisters is severely decreased as well. Agreement between analytical and numerical solutions is better than in a constant environment. Normals go extinct nearly exclusively during the catastrophe, which completely determines the fast part of the trajectory, rendering the instanton very simple. In turn, back-reaction of persisters becomes less important, and replacing the fast driving terms by step functions on the slow time scale becomes more accurate. These are the main approximations of the zeroth-order theory, hence the predictions improve with increasing catastrophe duration. We also confirmed that in all projections, the theory becomes more accurate with decreasing ε. At the same time, the "wiggles" identified in Sec. III C become less pronounced. Both tendencies go hand in hand, as both are based on reducing back-reaction. In Fig. 5, we compare the action (37) with numerical results. Even for moderately rare switching ($\varepsilon = 0.1$), the analytical prediction is reasonably accurate, and improving with increasing catastrophe duration.

We summarize the effect of the catastrophe in the leading order of rare switching. Independent of its duration, the strength of the catastrophe is set by the (normalized) death rate of normals. Normals decay exponentially on the very fast scale $t \sim 1$, responsible for the major part of phase space motion (unless $t_c \ll 1$). Persisters are affected indirectly via switching between the two populations. For a short catastrophe, $T_c' \ll 1$, the effect on Y_0 and P_{Y0} is negligible; switching hardly occurs during T_c', and the slow dynamics cannot resolve the difference in driving. Therefore only the normal action is reduced, and persisters are perfectly buffered against the catastrophe. Note that the time $t_c \ll 1/(\delta\varepsilon)$ can be much longer than the typical lifetime of an individual normal ~ 1. If the catastrophe is long enough to be seen on the slow scale, $T_c' \sim 1$ or larger, switching has an effect. While persisters still cannot resolve the accelerated extinction of normals, they trace the delay between X_0 and P_{X0} in the instanton. On the slow switching time scale it appears far shorter, forming a buffer that mitigates the catastrophe. The structure of the EPI (38) is thus based on the separation between the time scale of the catastrophe effect (strength), and the far longer time scale of persister dynamics. The catastrophe affects both populations, but acting on normals, its duration is measured on the very fast scale $t \sim 1$ of the death rate [action scales $\sim \exp(-t_c)$]; acting on persisters it is measured on the slow scale $T' \sim 1$ of switching back to normals [$\sim \exp(-T_c')$]. The crossover shows prominently in Fig. 5.

IV. WHY ARE SWITCHING RATES SMALL IN NATURE?

So far, we have considered small switching rates between the normal and persister states, $\varepsilon \ll 1$. The corresponding time-scale separation was the basis of our qualitative explanation and analytical treatment of the system's dynamics. We also numerically examined what happens at $\varepsilon \sim 1$ or larger. We still consider the system described by the Hamiltonians (11) and (26), respectively, as motivated at the end of Sec. III B. The instanton and the associated action are again obtained as detailed in Secs. II C and III C.

We found that both with and without catastrophe, instanton trajectories are qualitatively similar to the $\varepsilon \ll 1$ case even

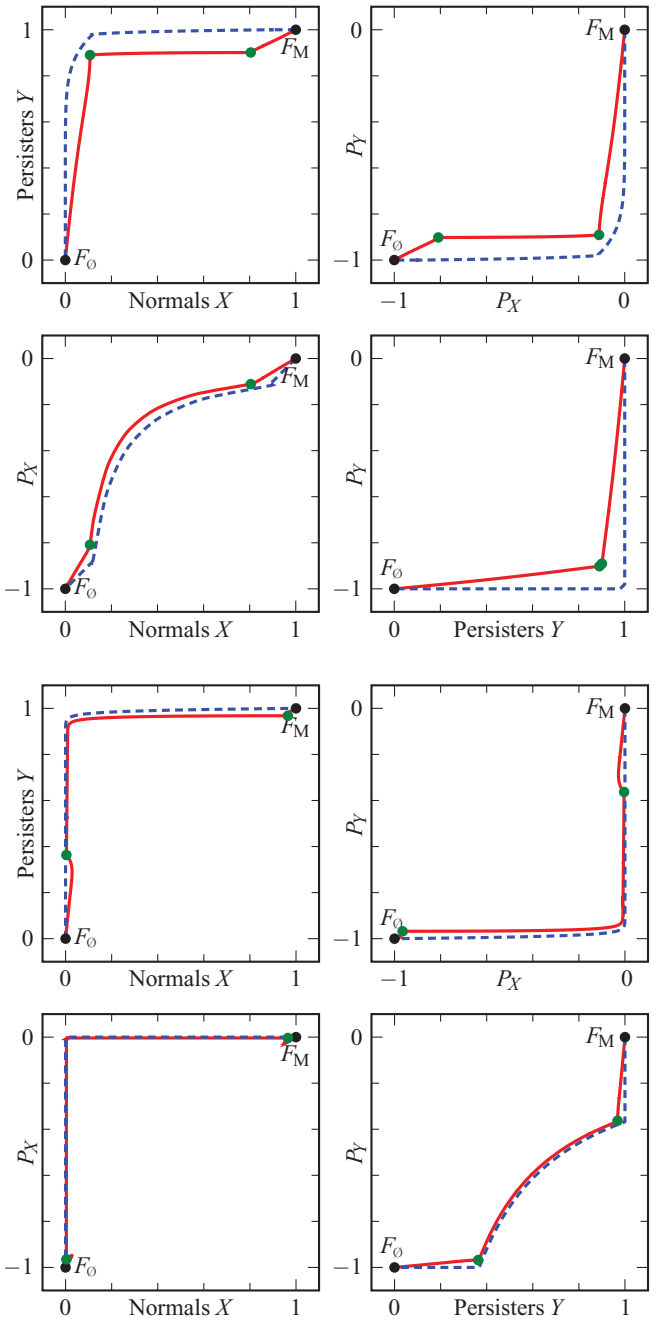

FIG. 4. (Color online) Instanton for $\Gamma = 1$, $\varepsilon = 0.1$, and $\delta = 0.1$, with a catastrophe of duration $T_c = 0.2$ (top) and 10 (bottom), respectively. Prediction by theory (dashed blue) and numerical solution (solid red). Green dots mark both the start and the end of the catastrophe.

when $\varepsilon = 1$. Further increasing ε "locks" persisters ever stronger to the dynamics of normals, see Eqs. (13) and (28). For very large ε, $P_Y \simeq P_X$ and $Y \simeq \Gamma X$ with only small deviations. Moreover, persisters still increase the action compared with a normals-only system of the same carrying capacity. We examined the action as a function of varying switching rate ε and catastrophe duration t_c (N, δ, Γ, and hence the carrying capacity K, being fixed). As expected, for given ε, the action decreases with increasing catastrophe duration t_c, and this decrease becomes stronger for larger switching

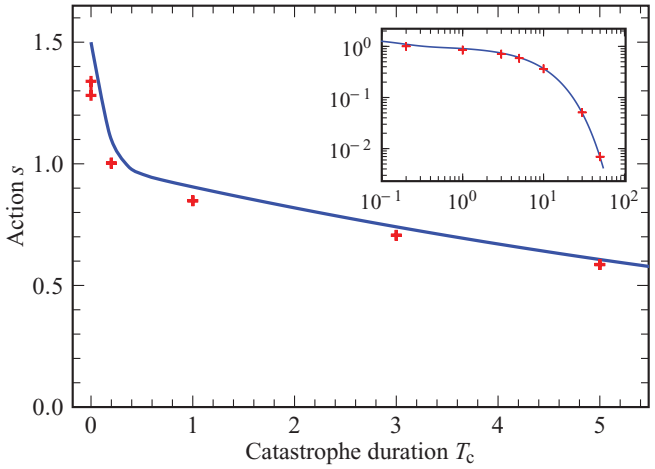

FIG. 5. (Color online) Action s vs catastrophe duration T_c, analytical (37) (solid blue line) and numerical result (red pluses), for $\Gamma = 1$, $\varepsilon = 0.1$, and $\delta = 0.1$. Error bars span the results obtained using the original action expression and its integrated-by-parts counterpart.

rate ε; the more frequent the switching is, the less insurance against extinction persisters provide. For given t_c, the action decreases with increasing switching rate ε, and this decrease becomes stronger for longer catastrophe duration: persisters are especially beneficial in the face of a catastrophe.

For very frequent switching, there is a new time-scale separation, which permits an analytical treatment. Consider the case $\delta \ll 1 \ll \alpha, \beta, (\varepsilon, \varepsilon\Gamma \gg 1/\delta)$, such that switching is frequent compared with the normal dynamics even during the catastrophe. In both the favorable [see Eqs. (13)] and catastrophic [see Eqs. (28)] environments, we have

$$Y = \Gamma X - \frac{1}{\varepsilon}\frac{dY}{dT}, \qquad P_Y = P_X + \frac{1}{\varepsilon}\frac{dP_Y}{dT}. \tag{40}$$

For large ε, the second term is a small correction, and we obtain

$$Y = \Gamma X - \frac{\Gamma}{\varepsilon}\frac{dX}{dT} + \ldots, \qquad P_Y = P_X + \frac{1}{\varepsilon}\frac{dP_X}{dT} + \ldots. \tag{41}$$

Inserting into the normal equations of motion yields, in the leading order in $1/\varepsilon$, the normals-only equations, but with a rescaled time $\tilde{T} = T/(1+\Gamma)$:

$$\frac{dX}{d\tilde{T}} = X(2P_X - X + 1), \qquad \frac{dP_X}{d\tilde{T}} = -P_X(P_X - 2X + 1) \tag{42}$$

in a favorable and

$$\frac{dX}{d\tilde{T}} = -\frac{X}{\delta} + XP_X, \qquad \frac{dP_X}{d\tilde{T}} = \frac{P_X}{\delta} - \frac{P_X^2}{2} \tag{43}$$

in a catastrophic environment. As in Sec. III D, from this we get X as of Eq. (31), only with the substitutions $T_{(c)} \to T_{(c)}/(1+\Gamma)$, and likewise for the momentum P_X and the energy $h_{X,c}$. Y and P_Y are given by Eq. (41).

Calculating the action along this instanton, first note that the switching term in the Hamiltonian is $h_{Y(,c)} \sim 1/\varepsilon$ at all

times. Second, the corrections in Eq. (41) do not contribute to the leading-order action, which becomes

$$s_c \simeq \int_{\text{pre/post}} (P_X\,dX + P_Y\,dY) + \int_{\text{cat.}} (P_X\,dX + P_Y\,dY - h_{X,c}\,dT)$$
$$\simeq (1+\Gamma) \int_{\text{pre/post}} P_X\,dX + (1+\Gamma) \int_{\text{cat.}} P_X\,dX - h_{X,c} T_c. \tag{44}$$

The second and third terms cancel; factoring out $(1+\Gamma)$, both contributions are the same as in the rare-switching case, only with the above rescaling applied to all times. The first integral is also known from the rare-switching case, where it coincided with the total action contributed by normals. Applying the time rescaling, the action thus becomes

$$s_c \simeq \frac{1+\Gamma}{1+e^{T_c/[\delta(1+\Gamma)]}}. \tag{45}$$

We confirmed (for $\varepsilon = 100$ and various values of Γ and t_c) that this agrees excellently with the action found numerically as described at the beginning of this section. This result is easily interpreted; very frequent switching effectively "mixes" the two subpopulations, as they rapidly switch back and forth. Compared with a normals-only system, the factor $1+\Gamma$ in the numerator describes the increased size of the combined population. A more subtle effect is the reduction, by the same factor $1+\Gamma$, of the effective duration of the catastrophe. This reduction accounts for the lag still gained by switching to the persister state.

In a favorable environment ($t_c = 0$), persisters switching frequently do not provide any benefit compared with a normals-only system of the same carrying capacity. With a catastrophic phase, however, we obtain

$$\frac{\Delta\mathcal{P}_{0,0}}{\Delta\mathcal{P}_{0,0}^{\text{1d}}} = \exp\left[-K\delta\left(\frac{1}{1+e^{t_c/(1+\Gamma)}} - \frac{1}{1+e^{t_c}}\right)\right]. \tag{46}$$

This is still a substantial benefit, although much less (for Γ not too large) than that for rarely switching persisters, see Eq. (39). Note that here Γ only appears in the effective catastrophe duration, not as the persister fraction.

Persisters are thus most valuable when stochastic switching is relatively rare. The fact that rare switching dominates in nature can be attributed to an evolutionary process.

V. DISCUSSION AND CONCLUSIONS

We have used a simple two-population model of normals and persisters to show that (and how) persisters exponentially decrease the extinction risk of an established bacterial population. We have compared the two-population system of normals and persisters to a normals-only system starting from the same total population size. Already in a constant environment favorable for normals, it is beneficial to switch to the persister state: persisters contribute to the MTE exponentially more than normals since their extinction is delayed by first switching back. When the population is under stress—that we model as a catastrophe—the same buffering is effective, rendering persisters far less prone to extinction, so that they exponentially reduce the EPI due to the catastrophe. For catastrophes that are long compared with the lifetime of normals but short compared with the much longer switching time scale (from persisters to normals), the reduction factor is squared with respect to the MTE increase in a constant environment: persisters are even more valuable for the population if it faces a catastrophe.

In exponential-growth models, which focus on fitness, persisters are only a burden in a constant favorable environment. To explain their existence with an overall benefit one needs to invoke temporary adverse conditions. In contrast, we have shown that persisters are always beneficial as an insurance against the extinction of an established population, as measured by the increased MTE, or by the reduced EPI during a catastrophe, respectively. We have also shown that to provide the optimal benefit, switching to and from the persister state has to be rare compared with all other processes. In this sense, the extinction risk perspective presented here explains, in a natural and robust way, the existence of persister phenotypes in bacteria as well as the small switching rates from the normals to persisters and back.

Our main analytical results (23) and (39) advocate that switching back from persisters to normals should be rare compared with switching to the persister state, leading to the largest possible fraction of persisters in the metastable state (within the range where our theory applies). For a bacterial population optimized solely against extinction from the established state, this would be an intuitive strategy even in favorable conditions. During the growth stage, on the other hand, the population needs optimal fitness to establish itself. These two complementary strategies, optimizing two different quantities, are not incompatible. The extinction risk perspective explains the mere existence of persisters, already without invoking environmental variations. The switching rates themselves (fixing the metastable persister fraction) may be tuned by evolution to optimize the growth stage in a variable environment.

Our simple model neglects many features that can be biologically relevant. For example, in reality persisters have reduced but nonzero birth and death rates. Such a more realistic system still features time-scale separation; persisters are now directly affected by a catastrophe (e.g., inhibiting their births), but again on a much slower scale than normals. Therefore we expect a qualitatively similar behavior. Future work can attempt to account for the cost of switching to persisters, for example, via competition between persisters and normals. There are also many alternatives for the detailed dynamics during the catastrophe. For many of them, the WKB approximation to the master equation provides a viable theoretical framework for determining the long-time behavior of bacterial populations.

ACKNOWLEDGMENTS

We appreciate useful discussions with Nathalie Q. Balaban. This work was supported by the Minerva foundation (IL), by the Israel Science Foundation (Grant No. 408/08), and by the US-Israel Binational Science Foundation (Grant No. 2008075).

[1] J. W. Bigger, Lancet **244**, 497 (1944).
[2] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, Science **305**, 1622 (2004).
[3] M. Lachmann and E. Jablonka, J. Theor. Biol. **181**, 1 (1996).
[4] F. Menu, J. Roebuck, and M. Viala, Am. Nat. **155**, 724 (2000).
[5] M. Thattai and A. van Oudenaarden, Genetics **167**, 523 (2004).
[6] D. M. Wolf, V. V. Vaziranib, and A. P. Arkin, J. Theor. Biol. **234**, 227 (2005).
[7] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler, Genetics **169**, 1807 (2005).
[8] E. Kussell and S. Leibler, Science **309**, 2075 (2005).
[9] M. J. Gander, C. Mazza, and H. Rummler, J. Math. Biol. **55**, 249 (2007).
[10] M. Acar, J. T. Mettetal, and A. van Oudenaarden, Nat. Genet. **40**, 471 (2008).
[11] P. Visco, R. J. Allen, S. N. Majumdar, and M. R. Evans, Biophys. J. **98**, 1099 (2010).
[12] J. D. Murray, *Mathematical Biology*, 3rd ed. (Springer, 2002), Vol. 1.
[13] O. Ovaskainen and B. Meerson, Trends in Ecology and Evolution **25**, 643 (2010).
[14] M. Assaf, A. Kamenev, and B. Meerson, Phys. Rev. E **79**, 011127 (2009).
[15] G. H. Weiss and M. Dishon, Math. Biosci. **11**, 261 (1971).
[16] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, J. Chem. Phys. **100**, 5735 (1994).
[17] O. A. van Herwaarden and J. Grasman, J. Math. Biol. **33**, 581 (1995).
[18] V. Elgart and A. Kamenev, Phys. Rev. E **70**, 041106 (2004).
[19] M. Assaf and B. Meerson, Phys. Rev. E **75**, 031122 (2007).
[20] A. Kamenev and B. Meerson, Phys. Rev. E **77**, 061107 (2008).
[21] M. I. Dykman, I. B. Schwartz, and A. S. Landsman, Phys. Rev. Lett. **101**, 078101 (2008).
[22] M. Khasin and M. I. Dykman, Phys. Rev. Lett. **103**, 068101 (2009).
[23] M. Khasin, B. Meerson, and P. V. Sasorov, Phys. Rev. E **81**, 031126 (2010).
[24] M. Khasin, M. I. Dykman, and B. Meerson, Phys. Rev. E **81**, 051925 (2010).
[25] M. Assaf and B. Meerson, Phys. Rev. E **81**, 021116 (2010).
[26] R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973).
[27] H. Gang, Phys. Rev. A **36**, 5782 (1987).
[28] C. S. Peters, M. Mangel, and R. F. Costantino, Bull. Math. Biol. **51**, 625 (1989).
[29] V. Elgart and A. Kamenev, Phys. Rev. E **74**, 041101 (2006).
[30] D. A. Kessler and N. M. Shnerb, J. Stat. Phys. **127**, 861 (2007).
[31] A. I. Chernykh and M. G. Stepanov, Phys. Rev. E **64**, 026306 (2001).